ICASE

THEORY OF SPECTRAL METHODS

FOR MIXED INITIAL-BOUNDARY VALUE PROBLEMS

PART II

(NASA-CR-185741) THEORY OF SPECTRAL METHODS FOR MIXED INITIAL-BOUNDARY VALUE PROBLEMS, PART 2 (ICASE) 183 p

N89-71432

Unclas

00/64 0224334

DAVID GOTTLIEB

STEVEN A. ORSZAG

Report Number 77-11

July 21, 1977

INSTITUTE FOR COMPUTER APPLICATIONS IN SCIENCE AND ENGINEERING

NASA Langley Research Center, Hampton, Virginia

Operated by the

UNIVERSITIES SPACE USRA RESEARCH ASSOCIATION

THEORY OF SPECTRAL METHODS
FOR MIXED INITIAL-BOUNDARY VALUE PROBLEMS
PART II

David Gottlieb

Steven A. Orszag*

* Department of Mathematics, Massachusetts Institute of Technology

This paper was prepared as a result of work performed under NASA Contract Number NAS1-14101 while the first author was in residence at ICASE, NASA Langley Research Center, Hampton, VA 23665. The work performed by the second author was supported by the Office of Naval Research and the National Science Foundation.

6. Spectral Methods Using Fourier Series

Fourier series are appropriate to solve problems with periodic boundary conditions. With periodic boundary conditions, a stable spectral method based on Fourier series is usually accurate and efficient. On the other hand, when Fourier series are used to solve non-periodic problems (including problems having periodic initial conditions but whose evolution operators violate periodicity), stability is not enough to ensure convergence to the true solution of the problem. An example of the latter effect was given in Example 1.3. In this section, we investigate the stability and convergence of spectral methods based on Fourier series.

Example 6.1: Constant-coefficient hyperbolic equation with periodic boundary conditions

Consider the one dimensional wave equation

$$u_t + u_x = 0 \qquad (0 \le x \le 1), \tag{6.1}$$

$$u(x,0) = f(x),$$

with periodic boundary conditions

$$u(0,t) = u(1,t) .$$

Since collocation, Galerkin and tau methods are identical in the absence of essential boundary conditions (see Sec. 2), let us analyze the Fourier-collocation or pseudospectral method.

We introduce the collocation points $x_n = n/2N$ $(n = 0,\ldots,2N-1)$ and the vector notation $\vec{u} = (u_0,\ldots,u_{2N-1})$ where $u_n = u(x_n)$. The collocation equations that approximate (6.1) can be written as

$$\frac{\partial \vec{u}}{\partial t} = C^{-1} D C \vec{u} , \tag{6.2}$$

where C and D are $2N \times 2N$ matrices whose entries are

$$C_{k\ell} = \frac{1}{\sqrt{2N}} \exp[-2\pi i (k-N) x_\ell] , \tag{6.3a}$$

$$D_{k\ell} = -2\pi i\, k'\, \delta_{k\ell} , \tag{6.3b}$$

where $k' = k-N$ $(1 \le k \le 2N-1)$ and $k' = 0$ if $k = 0$. A simple derivation of (6.2) is obtained by observing that $C\vec{u}$ gives the Fourier coefficients of the collocation projection Pu of $u(x)$. Thus, $DC\vec{u}$ are the Fourier coefficients of $-\frac{\partial}{\partial x} Pu$ and, finally, $C^{-1} DC\vec{u}$ gives the collocation projection of $-\frac{\partial}{\partial x}$ Pu which is $-P\frac{\partial}{\partial x}$ Pu. The matrix C is a unitary matrix so $C^* = C^{-1}$, and the matrix D is skew-Hermitian so $D^* = -D$. Therefore, $C^{-1}DC$ is skew-Hermitian so that

$$\|\exp[C^{-1}D\,C]t\| = 1 . \tag{6.4}$$

This proves that the Fourier-collocation method is stable for (6.1). The results of this example can be generalized to a general system of constant coefficient hyperbolic equations.

Example 6.2: <u>Variable-coefficient hyperbolic equation with periodic boundary conditions</u>

Consider the system of equations

$$u_t + A(x)u_x = 0 \qquad 0 \leq x \leq 1$$

with periodic boundary conditions $u(0,t) = u(1,t)$ and periodic inhomogenity: $A(x) = A(x+1)$ for all x. Here $u(x)$ is a vector of m components and $A(x)$ is an $m \times m$ matrix. If we assume that $A(x)$ is a symmetric matrix and that

$$\frac{\partial A}{\partial x} \leq \alpha I \tag{6.5}$$

for some finite α, then the Fourier-Galerkin method is stable. To show this, we denote by u_N the N-term Fourier-Galerkin approximation of u. Using (2.6.7) and integration by parts, we obtain

$$\frac{d}{dt}\int_0^1 u_N^* u_N dx = \int_0^1 u_N^*(A+A^*)_x u_N dx \leq 2\alpha \int_0^1 u_N^* u_N dx .$$

Therefore,

$$\int_0^1 u_N^*(x,t)\,u_N(x,t)\,dx \leq e^{2\alpha t}\int_0^1 u_N^*(x,0)\,u_N(x,0)\,dx$$

which proves stability.

Condition (6.5) is not sufficient to ensure stability for the collocation method. Consider the scalar equation $(m = 1)$

$$\begin{aligned} u_t &= r(x)\,u_x \qquad 0 \leq x \leq 1 \\ u(0,t) &= u(1,t) . \end{aligned} \tag{6.6}$$

If we impose the additional restriction that $r(x)$ is non-zero within $0 \leq x \leq 1$, then we can prove that the collocation method is stable. To do this, we show that exp(RC*DCt) is stable where C and D are given by (6.3) and R is the matrix with entries

$$R_{ij} = r(x_i)\ \delta_{ij} .$$

The matrix R^{-1} can be identified with the Liapounov matrix H_N invoked in (5.7) and, therefore, the method is stable:

$$R^{-1}(RC^*DC) + (C^*D^*CR^*)\, R^{-1} = 0.$$

In fact, following the argument leading to (5.11) gives

$$\|\exp(RC^*DCt)\|^2 \le \|R\|\,\|R^{-1}\| \le \max_{0\le x\le 1} |r(x)| / \min_{0\le x\le 1} |r(x)|,$$

proving stability for $N \to \infty$.

If $r(x)$ has a zero within $0<x<1$, collocation with Fourier series may lead to instability. For example, if $N = 2$, the eigenvalues of RC^*DC are $0, 0, \pm\sqrt{-(r_0+r_2)(r_1+r_3)}$ where $r_i = r(x_i)$, so there are growing modes if $(r_0+r_2)(r_1+r_3) < 0$. In some cases, these modes may have large growth rates. One way to limit the growth rate of these modes is to rewrite (6.6) as

$$u_t + \frac{1}{2}(r(x)u)_x + \frac{1}{2} r(x) u_x - \frac{1}{2}\frac{dr(x)}{dx} u = 0 \qquad (6.7)$$

Now Fourier-collocation gives the matrix equation

$$\vec{u}_t + (\tfrac{1}{2}C^*DCR + \tfrac{1}{2}RC^*DC - Q)\vec{u} = 0$$

where $Q_{k\ell} = -\frac{1}{2} r'(x_k)\delta_{k\ell}$. The first two matrices on the right side add up to a skew-Hermitian matrix. Also, if (6.5) holds for $r(x)$ then $Q \leq \frac{1}{2}\alpha I$. Therefore, we obtain the inequality

$$\frac{d}{dt} |\vec{u}|^2 \leq \alpha |\vec{u}|^2 .$$

Thus, we see it is possible to bound a priori the growth of modes in the Fourier-collocation method for variable coefficient problems with periodic boundary conditions.

On the other hand, for problems with non-periodic boundary conditions, Fourier-spectral methods can produce wrong solutions even when they are stable. This is illustrated by Example 1.3 which we now study more carefully.

Example 6.3: Hyperbolic equation with non-periodic boundary conditions

Consider the problem (1.7):

$$\frac{\partial u(x,t)}{\partial t} + \frac{\partial u(x,t)}{\partial x} = x + t \qquad (0 < x < \pi \quad t > 0)$$

$$u(0,t) = 0 \qquad (t \geq 0) \qquad (6.8)$$

$$u(x,0) = 0 \qquad (0 \leq x \leq \pi)$$

The solution is

$$u(x,t) = xt \ .$$

If we attempt to solve (6.8) by Fourier sine series using the Galerkin procedure we obtain

$$u_N = \sum_{n=1}^{N} a_n \sin nx \tag{6.9}$$

$$\frac{da_n}{dt} = -\frac{4}{\pi} \sum_{\substack{m=1 \\ m+n \text{ odd}}}^{N} \frac{nm}{n^2-m^2} a_m - \frac{2}{n}(-1)^n + \frac{4}{\pi n} t\, e_n \tag{6.10}$$

where $e_n = 0$ if n is even and $e_n = 1$ if n is odd.

It is easy to verify that the above approximation is stable. If we write (6.10) in the form

$$\frac{d\vec{a}}{dt} = A_N \vec{a} + \vec{f}$$

where $\vec{a} = (a_1,\ldots,a_N)$, $\vec{f} = (f_1,\ldots,f_N)$, $f_n = \frac{2}{n}\left[(-1)^{n+1} + 2te_n/\pi\right]$, then

$$A_N + A_N^* = 0 \ .$$

Thus, $||\exp(A_N t)|| = 1$ for all N and t.

In Figs. 6.1-6.4 we plot the solution of (6.9-10) at $t = 1$ for $N = 25, 50, 75, 100$. It is apparent that $u_N(x,1)$ does not converge to the exact solution xt at $t = 1$ as $N \to \infty$. Instead, u_N for N even appears to be converging as

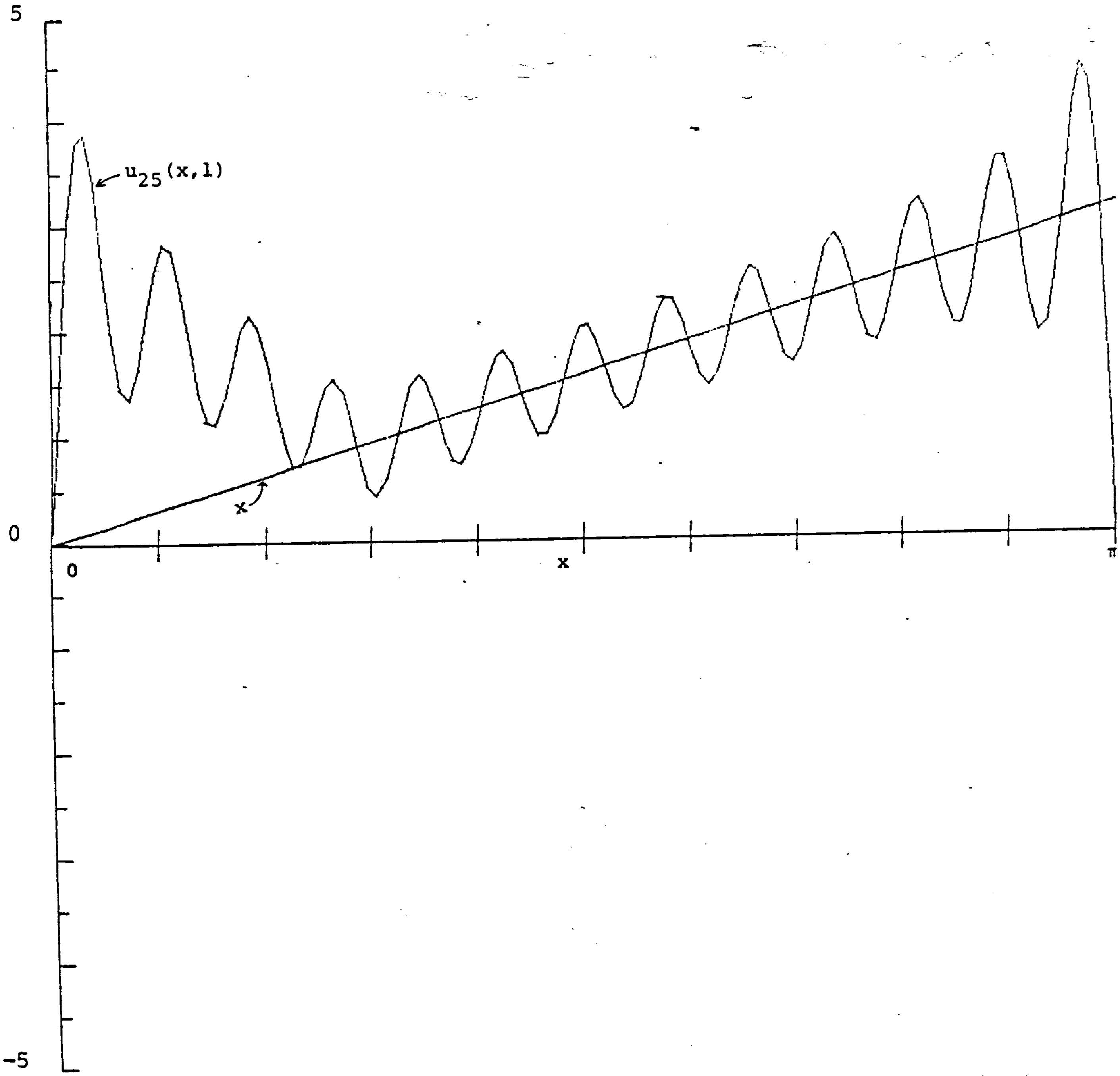


Fig. 6.1. A plot of $u_N(x,t)$ vs x for $N=25$ and $t=1$ where $u_N(x,t)$ is determined by numerical integration of (6.9-10) with negligible time-differencing errors. A plot of the exact solution xt at $t=1$ to (6.8) is also given. Observe the apparent divergence of $u_N(x,t)$ from the exact solution for $0 \leq x < t$ and the enhanced Gibbs phenomenon at $x=0,\pi$.

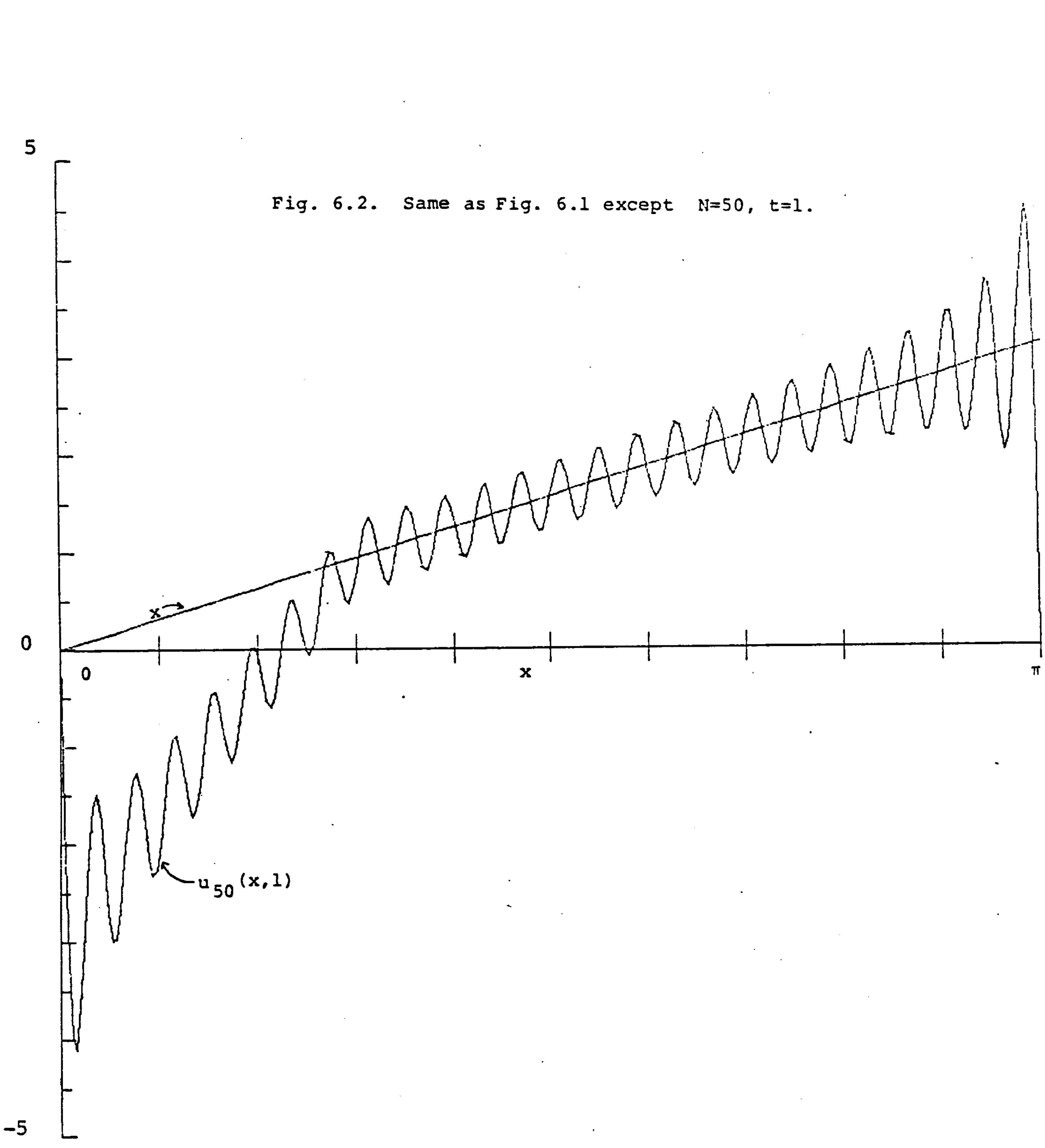
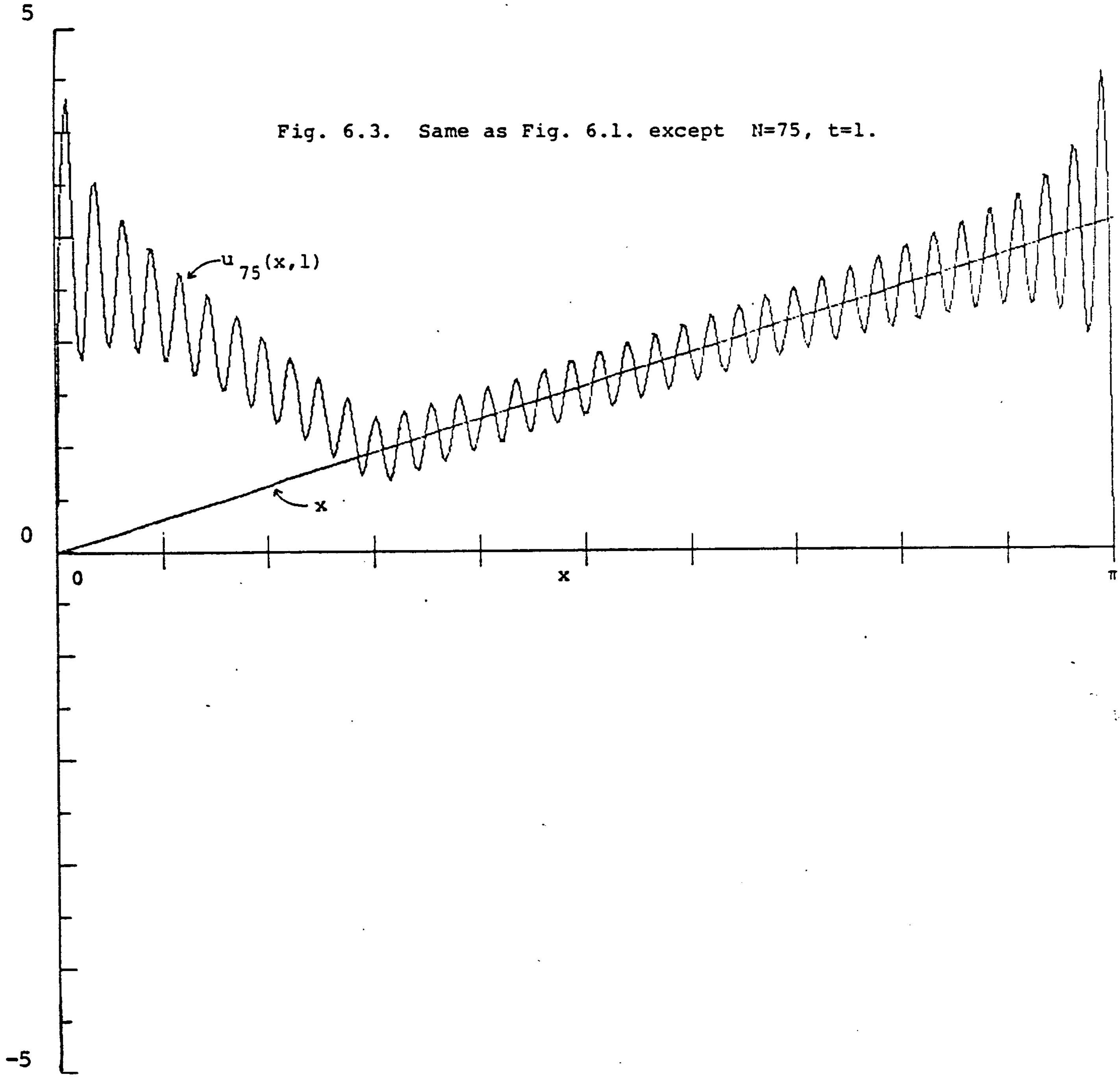


Fig. 6.2. Same as Fig. 6.1 except N=50, t=1.

Fig. 6.3. Same as Fig. 6.1. except N=75, t=1.



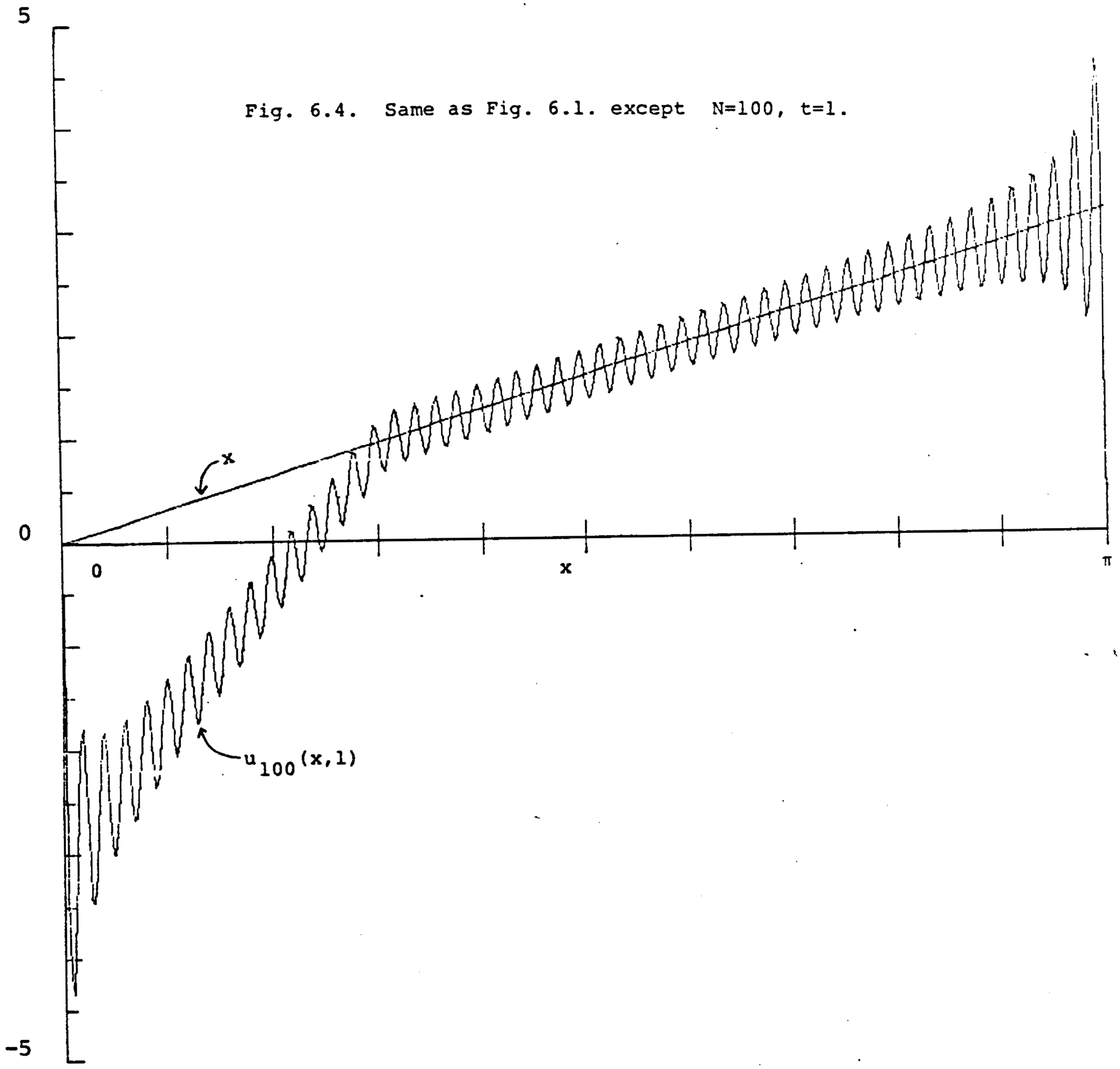


Fig. 6.4. Same as Fig. 6.1. except N=100, t=1.

$N \to \infty$ to the function

$$u_{even} = \begin{cases} xt & x \geq t \\ \pi(x-t)+xt & x < t\ , \end{cases} \tag{6.11}$$

for $t < \pi$, while u_N for N odd appears to converge to the function

$$u_{odd} = \begin{cases} xt & x \geq t \\ \pi(t-x)+xt & x < t \end{cases} \tag{6.12}$$

for $t < \pi$. The results plotted in Fig. 6.5 for $u_{100}(x,t=2)$ are also consistent with convergence to the wrong solution (6.11). Notice that the approximations $u_N(x,t)$ plotted in Figs. 6.1-5 all exhibit a large region of nonuniform convergence near $x = 0$ and $x = \pi$ and that the rate of convergence to the wrong solutions (6.11-12) in the interior of the interval $0<x<\pi$ is roughly like $1/\sqrt{N}$.

The origin of the divergence of (6.9-10) from the exact solution to (6.8) is not instability; rather, the divergence is due to inconsistency. Since $||\exp(A_N t)|| = 1$, the method is stable. To show that it is not consistent we estimate the truncation error in the L_2 norm,

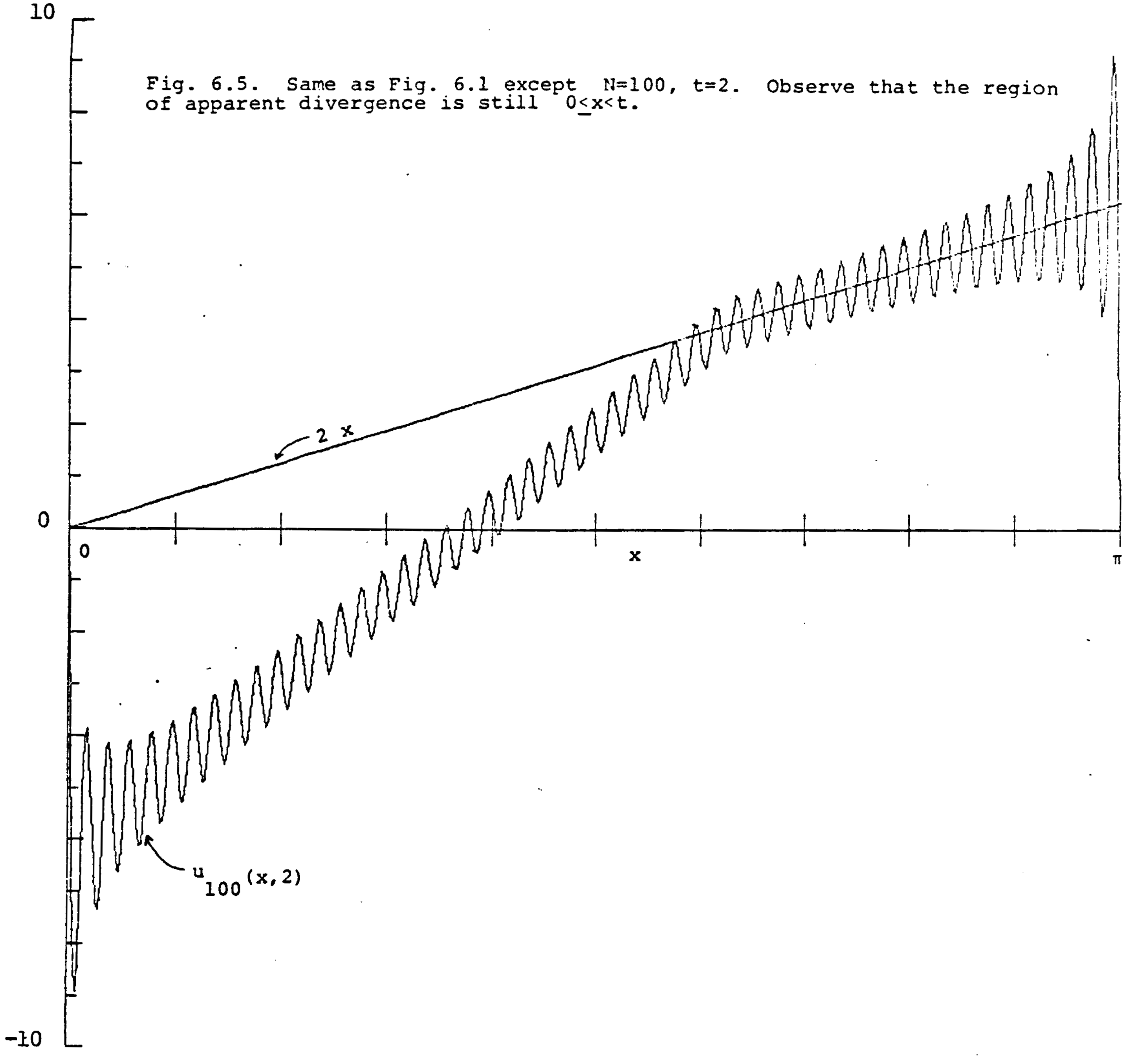


Fig. 6.5. Same as Fig. 6.1 except N=100, t=2. Observe that the region of apparent divergence is still $0 \leq x < t$.

$$\varepsilon_N = ||Lu - L_N u||,$$

for $u = xt$ where $L_N = P_N L P_N$ and P_N is the Galerkin projection operator and $L = -\partial/\partial x$. This error can be bounded from below by

$$\varepsilon_N = ||Lu - P_N Lu + P_N Lu - P_N L P_N u||$$

$$\geq ||P_N L(I-P_N)u|| - ||(I-P_N)Lu|| \quad .$$

However, $||(I-P_N)Lu|| \to 0$ (like $1/\sqrt{N}$) as $N \to \infty$ because this norm is just the error in the Fourier sine series expansion of $Lu = -\frac{\partial}{\partial x} xt = t$. Therefore, if we can show that $||P_N L(I-P_N)u||$ does not approach zero as $N \to \infty$ then (6.9-10) is not consistent.

To estimate $||P_N L(I-P_N)u||$ we proceed as follows. Since

$$(I-P_N)u = \sum_{n=N+1}^{\infty} a_n(t) \sin nx$$

we obtain

$$P_N L (I-P_N)u = \sum_{n=1}^{N} b_n(t) \sin nx$$

where

$$b_n(t) = -\frac{4}{\pi} \sum_{\substack{m=N+1 \\ m+n \text{ odd}}}^{\infty} \frac{nm}{n^2-m^2} a_m(t) \; .$$

Therefore, since the Fourier coefficients of u are given by $a_n(t) = 2(-1)^{n+1} t/n$,

$$\|P_N L(I-P_N)u\|^2 = \sum_{n=1}^{N} b_n^2$$

$$= \frac{64}{\pi^2} t^2 \sum_{n=1}^{N} \left(\sum_{\substack{m=N+1 \\ m+n \text{ odd}}}^{\infty} \frac{n}{n^2-m^2} \right)^2$$

$$\geq \frac{64t^2}{\pi^2} \sum_{n=1}^{N} \left(\sum_{\substack{m=N+1 \\ m+n \text{ odd}}}^{\infty} \frac{n}{m^2} \right)^2$$

$$\geq Ct^2 \sum_{n=1}^{N} \frac{n^2}{N^2} \geq C_1 t^2 N$$

for suitable constants C and C_1. This completes the proof that $||Lu - L_N u||$ does not approach zero as $N \to \infty$.

Blair Swartz (private communication, 1976) traces the inconsistency of (6.9-10) to the incompleteness of the set of functions $\{L(\sin n x) = - n \cos n x,\ n=1,2,\ldots\}$. This set of functions is made complete by augmenting it by the function 1. Whereas u may be well approximated by a function u_N of the form (6.9), Lu may not be well approximated by the function Lu_N. In fact, if $||Lu - Lu_N|| \to 0$ as $N \to \infty$, then

$$\int_0^\pi (Lu - Lu_N)\, dx \to 0 \ (N \to \infty).$$

Since

$$\int_0^\pi Lu_N = - \int_0^\pi \sum na_n \cos n x\, dx = 0,$$

Lu may be well approximated by Lu_N only if

$$0 = \int_0^\pi Lu\, dx = u(0) - u(\pi),$$

which is generally not true.

As shown in Figs. 6.1-5, $u_N(x,t)$ does not coverge to xt as $N \to \infty$. The analysis given above provides no clue to the fascinating way in which the method achieves this divergence. There is no indication of the 'error' wave $(-1)^N \pi(x-t)$ that appears in (6.11-12) and propagates with speed 1 across $0 < x < \pi$. It seems that the complete mathematical analysis of the divergence of (6.9-10) is difficult and we do not now have a justifiable argument to demonstrate convergence of u_N to u_{even} and u_{odd} given by (6.11-12) as $N \to \infty$ through even and odd values, respectively.

In the next example we will show that it is not simply the presence of boundary conditions but rather the non-periodic nature of the problem that causes the divergence of the Fourier-spectral methods.

Example 6.4 Non-periodic boundary-free problem

Consider the problem

$$\frac{\partial u}{\partial t} + \left(x-\frac{\pi}{2}\right) \frac{\partial u}{\partial x} = 0 \qquad (0 < x < \pi)$$
$$u(x,0) = f(x) \tag{6.13}$$

The problem is well posed without specifying any boundary condition. However, it is clear that the exact solution given by

$$u(x,t) = f\left(\frac{\pi}{2} + \left(x-\frac{\pi}{2}\right) e^{-t}\right) \tag{6.14}$$

is not periodic in x. Since $r(x) = x - \frac{\pi}{2}$ has a bounded derivative, it follows from Example 6.2 that Fourier-Galerkin approximation to (6.13) is stable. Nevertheless it is not convergent as shown by the results plotted in Figs. 6.6-8 for $f(x) = \sin x$ and $N = 5, 10$, and 20 retained terms in the Fourier sine series.

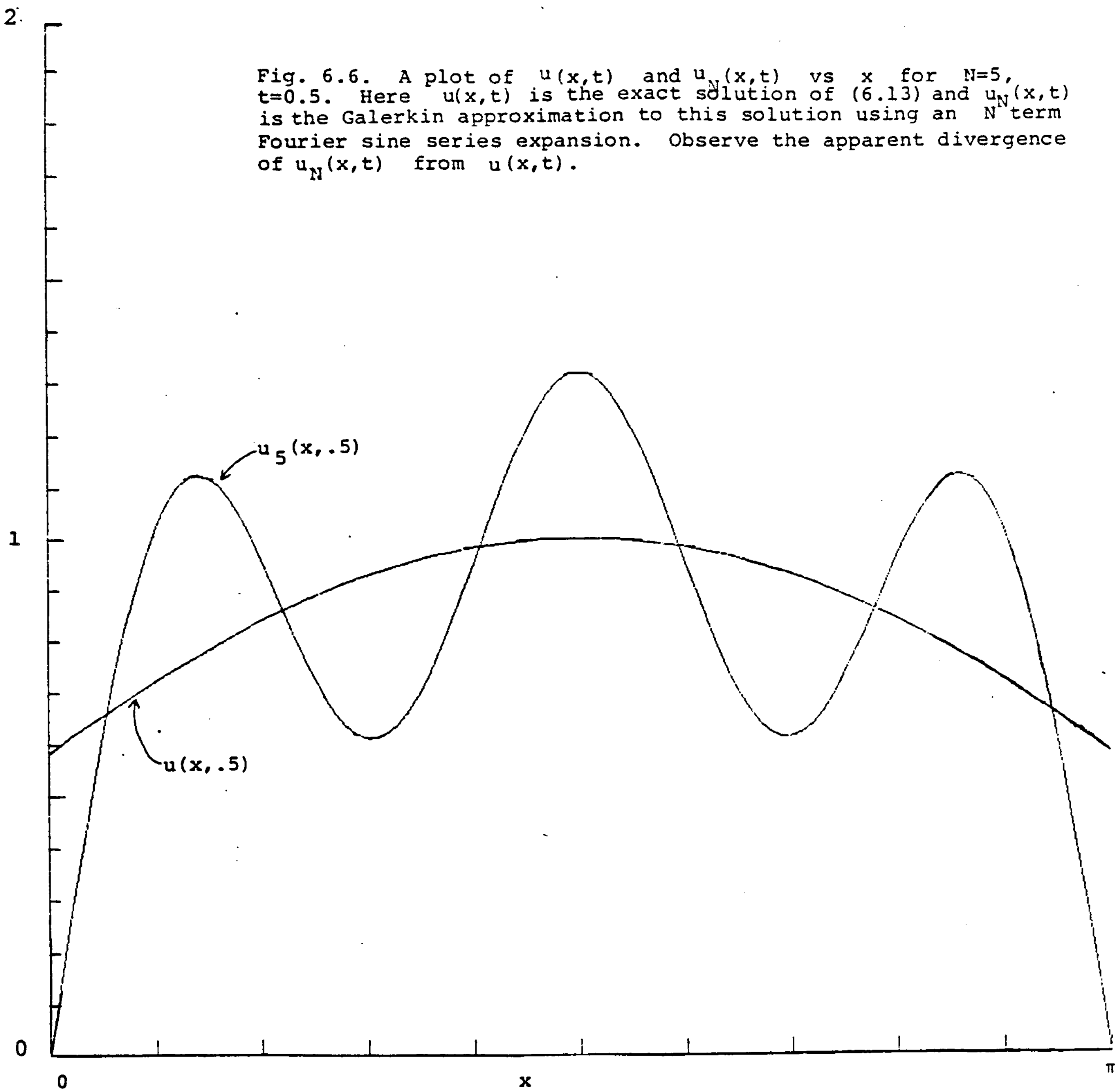


Fig. 6.6. A plot of $u(x,t)$ and $u_N(x,t)$ vs x for $N=5$, $t=0.5$. Here $u(x,t)$ is the exact solution of (6.13) and $u_N(x,t)$ is the Galerkin approximation to this solution using an N term Fourier sine series expansion. Observe the apparent divergence of $u_N(x,t)$ from $u(x,t)$.

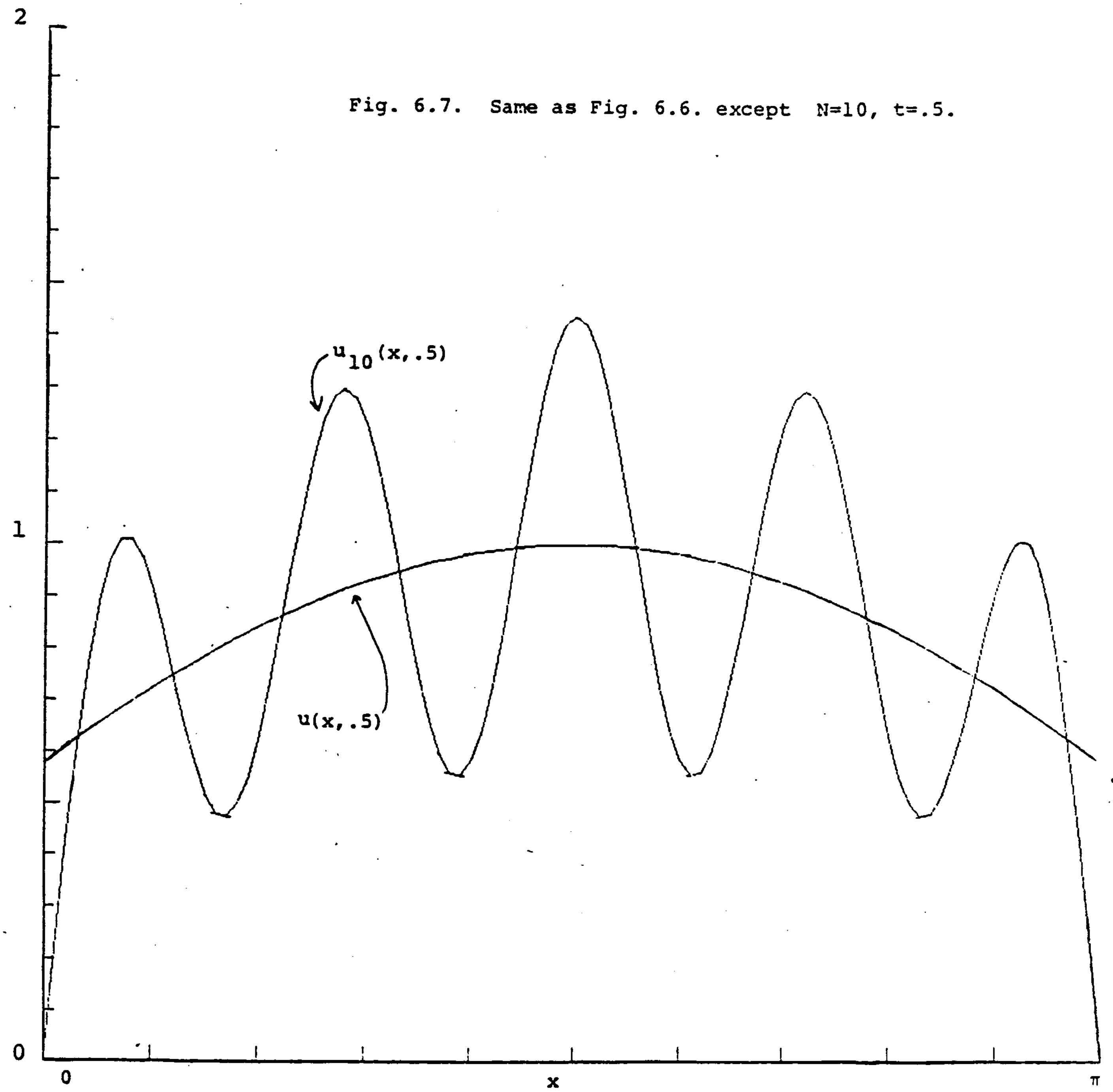


Fig. 6.7. Same as Fig. 6.6. except N=10, t=.5.

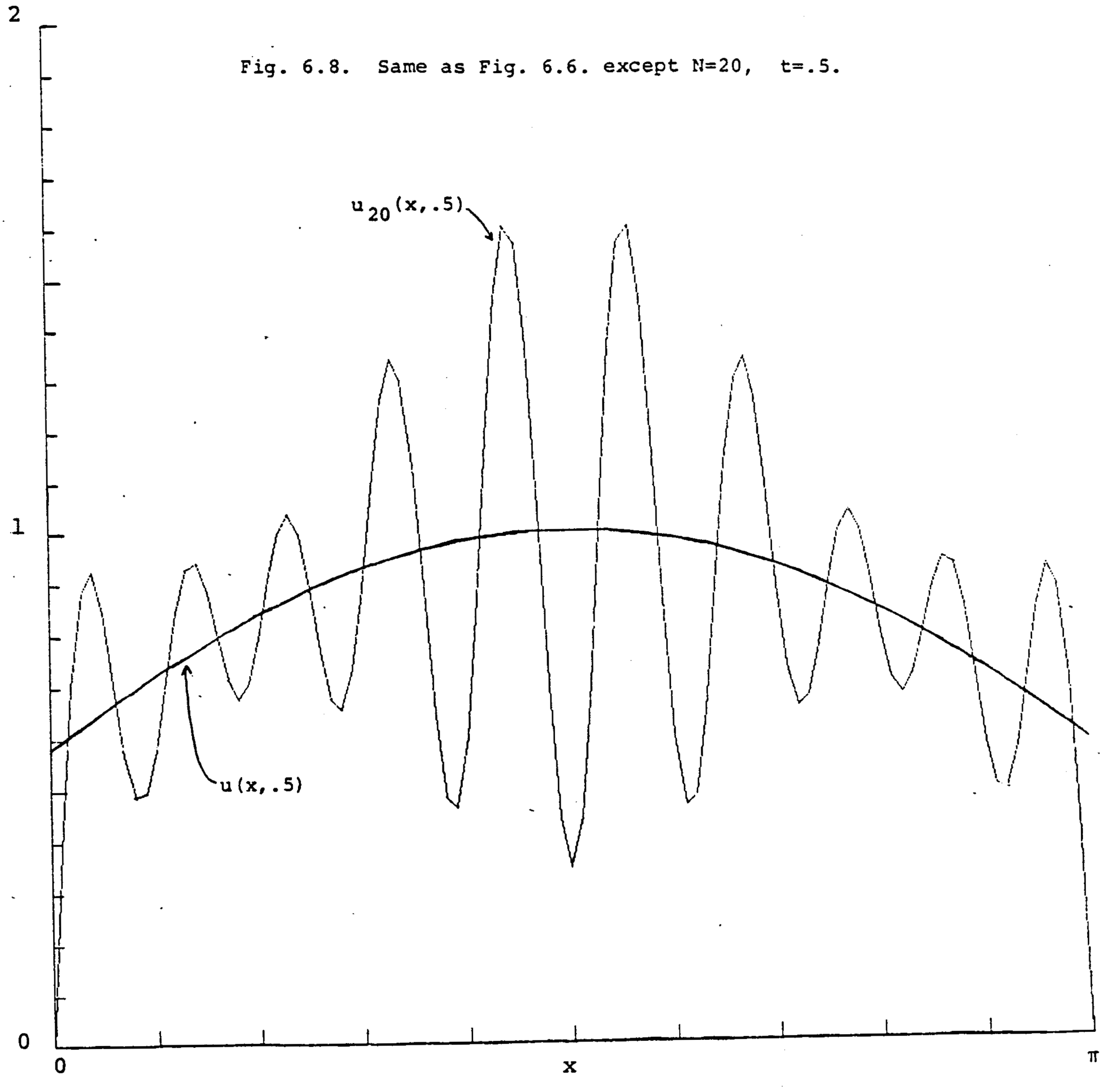


Fig. 6.8. Same as Fig. 6.6. except N=20, t=.5.

Polynomial Subtractions for Non-Periodic Problems

There is a method that can be used to ensure that Fourier series yield convergent results for non-periodic problems. The idea is to express the solution as the sum of a low-order polynomial and a Fourier series; the polynomial is chosen so that the Fourier series converges rapidly as suggested originally by Lanczos (1956,1966) . The method has been used by Orszag (1971c) and Wengle & Seinfeld (1977) to solve problems with non-periodic boundary conditions. We illustrate it here for the problem discussed in Example 6.4.

Example 6.5 Polynomial subtractions applied to Fourier series

The Fourier sine series expansion of the exact solution $u(x,t)$ to (6.13) converges slowly because, in general, $u(0,t) \neq 0$ and $u(\pi,t) \neq 0$. This slow convergence of the Fourier series of the exact solution implies that Galerkin approximation is inconsistent, as shown using the methods of Example 6.3. In order to avoid slow convergence or even divergence, we proceed as follows.

We seek the solution to (6.13) as the sum of a linear polynomial and a Fourier series:

$$u(x,t) = b(t)x + c(t)(\pi - x) + \sum_{n=1}^{\infty} a_n(t) \sin nx \qquad (6.15)$$

where $b(t)$ and $c(t)$ are chosen to ensure that $a_n(t) \to 0$ rapidly as $n \to \infty$. Substituting (6.15) into (6.13) gives

$$b'(t)x + c'(t)(\pi - x) + \sum_{n=1}^{\infty} a_n'(t)\sin nx = (\tfrac{\pi}{2} - x)[b(t) - c(t)]$$

$$+ \sum_{n=1}^{\infty} \hat{a}_n(t)\sin nx \tag{6.16}$$

where

$$\hat{a}_n(t) = \sum_{\substack{m=1 \\ n+m \text{ even} \\ n \neq m}}^{\infty} \frac{2nm}{n^2 - m^2} a_m + \frac{1}{2} a_n \tag{6.17}$$

are the Fourier sine coefficients of $(\frac{\pi}{2} - x)\frac{\partial}{\partial x}\Sigma a_n \sin nx$.

If we knew $u(0,t)$ and $u(\pi,t)$ we could set $b(t)=u(\pi,t)/\pi$ and $c(t)=u(0,t)/\pi$; with this choice, the Fourier sine series in (6.15) does not exhibit the Gibbs phenomenon and $a_n(t)=0(1/n^3)$ as $n\to\infty$. However, the boundary conditions on u are not known as part of the specifications of the problem (6.13). Therefore, we must solve for $b(t)$ and $c(t)$ directly from the differential equation.

Equating coefficients of $\sin nx$ in (6.16) gives

$$\frac{da_n}{dt} = [c'-b'+c-b]\frac{2}{n}(-1)^{n+1} + [b-c-2c']\frac{2}{n} e_n + \hat{a}_n \quad (n=1,\ldots) \tag{6.18}$$

where $e_n = 1$ if n is odd, 0 if n is even; here we use the Fourier sine series expansion of 1 and x:

$$1 = \frac{4}{\pi} \sum_{\substack{n=1 \\ n \text{ odd}}}^{\infty} \frac{\sin nx}{n} ,$$

$$x = 2 \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\sin nx}{n} .$$

Also, if $b(t)$ and $c(t)$ are chosen so that $a_n = O(1/n^3)$ as $n \to \infty$, then the Fourier series $\sum a_n \sin nx$ may be differentiated termwise so

$$\sum_{n=1}^{\infty} \hat{a}_n \sin nx = \left(\frac{\pi}{2} - x\right) \frac{\partial}{\partial x} \sum_{n=1}^{\infty} a_n \sin nx = \left(\frac{\pi}{2} - x\right) \sum_{n=1}^{\infty} n a_n \cos nx .$$

Therefore,

$$\lim_{x \to 0+} \sum_{n=1}^{\infty} \hat{a}_n \sin nx = \frac{\pi}{2} \sum_{n=1}^{\infty} n\, a_n ,$$

$$\lim_{x \to \pi-} \sum_{n=1}^{\infty} \hat{a}_n \sin nx = -\frac{\pi}{2} \sum_{n=1}^{\infty} (-1)^n n\, a_n .$$

Using these results and setting $x = \pi$ and $x = 0$ in (6.16) gives, respectively,

$$\frac{db}{dt} = \frac{1}{2} (c-b) - \frac{1}{2} \sum_{n=1}^{\infty} (-1)^n n a_n \tag{6.19}$$

$$\frac{dc}{dt} = \frac{1}{2} (b-c) + \frac{1}{2} \sum_{n=1}^{\infty} n\, a_n . \tag{6.20}$$

Galerkin approximation reproduces the equations (6.18-20) with $a_n = 0$ for $n = N+1, N+2, \ldots$.

The above derivation suggests, but does not prove, that $a_n(t) \to 0$ sufficiently rapidly as $n \to \infty$ so that inconsistency problems are avoided. The exact solution of (6.13), which satisfies (6.18-20) with $N = \infty$, does satisfy $a_n = 0(1/n^3)$ as $n \to \infty$. However, the Galerkin approximation with finite N does not yield such a rapidly converging result. In fact, estimates like those given in Example 6.3 show that

$$||Lv - L_N v|| = 0\left(\frac{1}{N^{3/2}}\right) \qquad (N \to \infty) \tag{6.21}$$

where v satisfies $v(0,t) = v(\pi,t) = 0$ and $L = (\frac{\pi}{2} - x)\frac{\partial}{\partial x}$. Since the Galerkin approximation (6.18) is stable (see Example 6.6), we expect that the errors in the Galerkin approximation (6.18-20) are of order $N^{-3/2}$ for fixed t.

The above prediction has been tested numerically. In Table 6.1 we list for various N the maximum errors in the approximation obtained by solving (6.18-20). A plot of the error $u_N(x,t) - u(x,t)$ vs x for $N = 30, 40$ at $t = .5$ is given in Fig. 6.9 - 10 .

In the next example, we prove that the method of polynomial subtraction used in Example 6.5 is stable.

Example 6.6. Proof of stability for polynomial subtractions

It is not obvious that the approximation (6.18-20) is stable. Fourier series approximation without polynomial subtractions are stable but not consistent (see Example 6.4). On the other hand,

Table 6.1

N	$\varepsilon_N = \max\lvert u_N(x,t=.5) - u(x, t=.5)\rvert$	$N^{3/2}\,\varepsilon_N$
5	4.19 (-3)	4.7 (-2)
10	2.13 (-3)	6.7 (-2)
15	1.13 (-3)	6.6 (-2)
20	8.28 (-4)	7.4 (-2)
25	5.76 (-4)	7.2 (-2)
30	4.70 (-4)	7.7 (-2)
35	3.64 (-4)	7.5 (-2)
40	3.13 (-4)	7.9 (-2)

Table 6.1. Errors in the polynomial-subtracted Fourier series approximation $u_N(x,t)$ given by (6.22) and (6.18-20) for the problem (6.13) with $f(x) = \sin x$ for $t=.5$. Observe that the errors appear to decrease as $N^{-3/2}$ as $N \to \infty$ in agreement with the estimate (6.21).

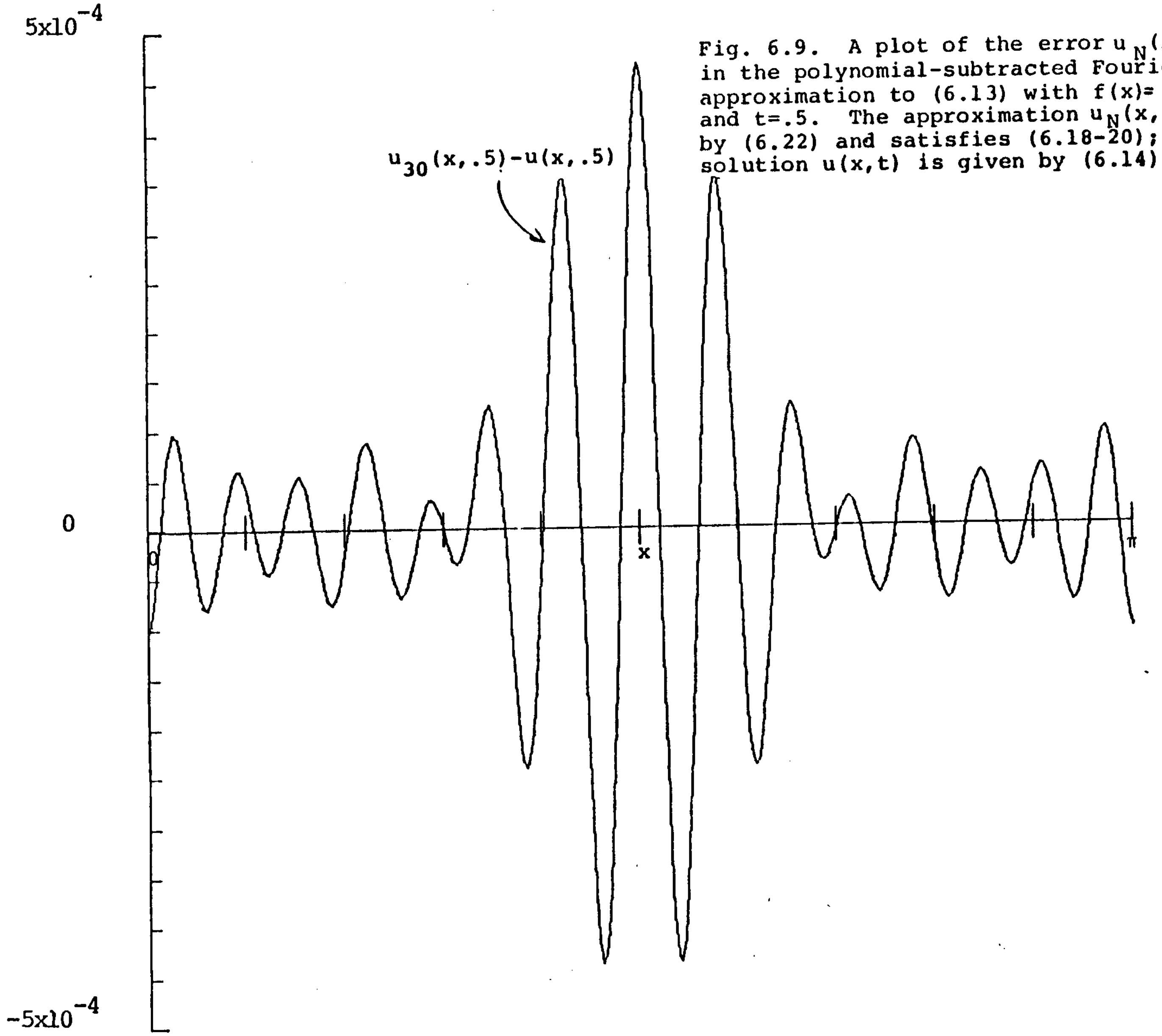


Fig. 6.9. A plot of the error $u_N(x,t)-u(x,t)$ in the polynomial-subtracted Fourier-series approximation to (6.13) with $f(x)=\sin x$, $N=30$ and $t=.5$. The approximation $u_N(x,t)$ is given by (6.22) and satisfies (6.18-20); the exact solution $u(x,t)$ is given by (6.14).

-26-

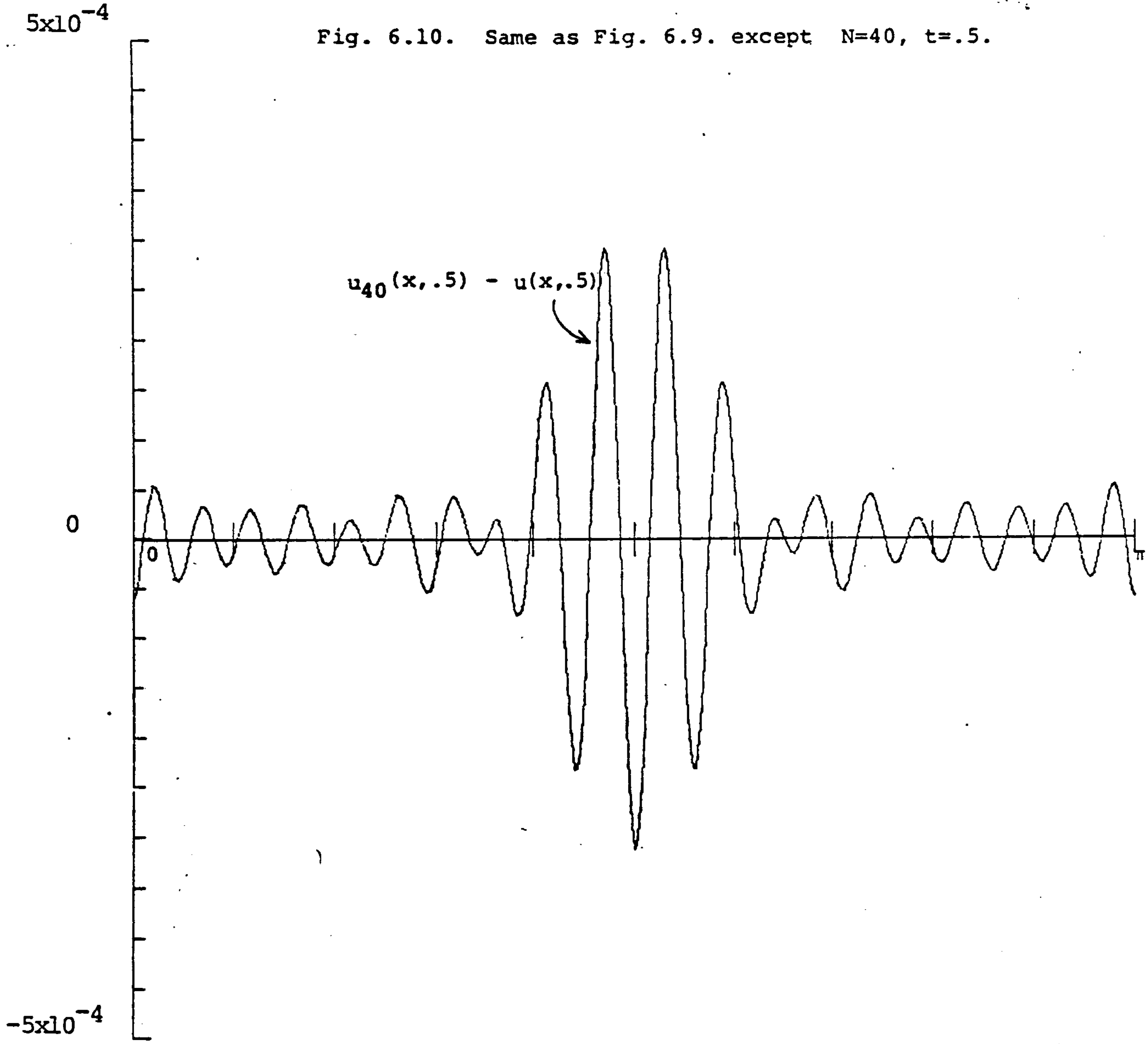


Fig. 6.10. Same as Fig. 6.9. except N=40, t=.5.

the approximations obtained by polynomial subtractions are consistent as shown by (6.21), but their stability remains to be shown.

To demonstrate stability of (6.18-20), we reformulate these equations in terms of $u_N(x,t)$ defined by

$$u_N(x,t) = b(t)x + c(t)(\pi - x) + \sum_{n=1}^{N} a_n(t) \sin nx . \qquad (6.22)$$

In terms of $u_N(x,t)$, (6.18) is equivalent to

$$\int_0^\pi \left[\frac{\partial u_N}{\partial t} + (x - \frac{\pi}{2}) \frac{\partial u_N}{\partial x}\right] \sin nx \; dx = 0 \quad (n=1,\ldots N), \qquad (6.23)$$

while (6.19-20) become, respectively,

$$\left[\frac{\partial u_N}{\partial t} + (x - \frac{\pi}{2}) \frac{\partial u_N}{\partial x}\right] \Bigg|_{x = \pi} = 0, \qquad (6.24)$$

$$\left[\frac{\partial u_N}{\partial t} + (x - \frac{\pi}{2}) \frac{\partial u_N}{\partial x}\right] \Bigg|_{x = 0} = 0, \qquad (6.25)$$

Multiplying (6.23) by $n^2 a_n$, suming from $n = 1$ to $n = N$, and noting that

$$\frac{\partial^2 u_N}{\partial x^2} = - \sum_{n=0}^{N} n^2 a_n \sin nx ,$$

we obtain

$$\int_0^\pi \left[\frac{\partial u_N}{\partial t} + (x-\frac{\pi}{2})\ \frac{\partial u_N}{\partial x}\right] \frac{\partial^2 u_N}{\partial x^2}\ dx = 0. \tag{6.26}$$

Integrating (6.26) once by parts and using (6.24-25), we obtain

$$\int_0^\pi \frac{\partial}{\partial x}\left[\frac{\partial u_N}{\partial t} + (x-\frac{\pi}{2})\ \frac{\partial u_N}{\partial x}\right] \frac{\partial u_N}{\partial x}\ dx = 0.$$

Therefore,

$$\frac{\partial}{\partial t}\int_0^\pi \left(\frac{\partial u_N}{\partial x}\right)^2 dx = -2\int_0^\pi \left(\frac{\partial u_N}{\partial x}\right)^2 dx - \int_0^\pi (x-\frac{\pi}{2})\ \frac{\partial}{\partial x}\left(\frac{\partial u_N}{\partial x}\right)^2 dx$$

Integrating the second integral on the right once by parts gives

$$\frac{\partial}{\partial t}\int_0^\pi \left(\frac{\partial u_N}{\partial x}\right)^2 dx = -\int_0^\pi \left(\frac{\partial u_N}{\partial x}\right)^2 dx - \frac{\pi}{2}\left[\left(\frac{\partial u_N}{\partial x}\right)^2\Bigg|_{x=\pi} + \left(\frac{\partial u_N}{\partial x}\right)^2\Bigg|_{x=0}\right],$$

so that

$$\frac{\partial}{\partial t}\int_0^\pi \left(\frac{\partial u_N}{\partial x}\right)^2 dx \leq -\int_0^\pi \left(\frac{\partial u_N}{\partial x}\right)^2 dx.$$

Thus, we obtain the stability estimate

$$\int_0^\pi \left[\frac{\partial u_N(x,t)}{\partial x}\right]^2 dx \leq e^{-t}\int_0^\pi \left[\frac{\partial u_N(x,0)}{\partial x}\right]^2 dx. \tag{6.27}$$

The bound (6.27) shows the stability of (6.18-20).

Examples 6.5-6 suggest that by subtracting polynomials of higher and higher degree from $u(x,t)$, the residual Fourier series can be made to converge faster and faster. Subtracting a linear polynomial as in (6.15) gives Fourier approximations with errors of order $N^{-3/2}$ as $N \to \infty$; subtracting a quadratic polynomial gives Fourier approximations with errors of order $N^{-7/2}$; and so on. In the limit we disperse entirely with Fourier series and obtain a rapidly converging polynomial approximation. The convergence theory of these polynomial spectral approximations is discussed in the next two sections.

7. Applications of Algebraic-Stability Analysis

The main result of Sec. 5 does not provide us with a systematic way of constructing the family H_N of Liapounov matrices necessary to prove algebraic stability. In general, these matrices are difficult to find. However, there are several problems for which they can be found directly from the differential equation.

It is very easy to construct Liapounov matrices for Galerkin approximations to

$$\frac{\partial u}{\partial t} = Lu$$

when L is a semi-bounded operator on the Hilbert space $\mathcal{H}$. We say that L is semi-bounded if

$$L + L^* \leq \alpha I \tag{7.1}$$

for some constant α, where L^* is the adjoint of L defined with respect to the Hilbert space inner product (,). If L is semi-bounded

$$\frac{d}{dt}(u,u) \leq \alpha(u,u) , \tag{7.2}$$

so

$$(u(t),u(t)) \leq e^{\alpha t}(u(0),u(0))$$

and the 'energy' $(u(t),u(t))$ grows at most exponentially with t.

If an energy estimate of the form (7.2) exists, then Galerkin approximation based on the Hilbert space inner product $(\cdot,\cdot)$ is stable (and, hence, algebraically stable). The Liapounov matrix H_N may be chosen to be the $N \times N$ identity matrix I_N. It follows from the Galerkin equations (2.6-7) with $f \equiv 0$ that

$$\frac{d}{dt}(u_N,u_N) = (u_N,[L+L^*]u_N) \le \alpha(u_N,u_N)$$

Thus,

$$(u_N(t),u_N(t)) \le e^{\alpha t}(u_N(0),u_N(0))$$

Since $u_N(t) = \exp(L_N t)u_N(0)$ for all $u_N(0)$, we obtain $\|\exp(L_N t)\| \le \exp(\frac{1}{2}\alpha t)$ so stability is proved. The reader is reminded that with stability established, the theory of Sects. 4 and 5 proves convergence for consistent schemes.

Example 7.1: Semi-bounded Galerkin approximations

The above construction establishes stability and thus convergence for a wide variety of Galerkin approximations. Among these stable Galerkin approximations are:

(i) Solution of any problem $u_t = Lu$ that is semi-bounded in $L_2(-1,1)$ by means of Legendre series. For example, $u_t + u_x = f(x,t)$ with $u(-1,t) = 0$ is stable (and convergent) when solved by Legendre-Galerkin approximation. For our argument to be complete it is necessary to verify that the Legendre-Galerkin approximation to this problem is consistent. This is done as follows.

We write

$$||Lu-P_N L P_N u|| \le ||(I-P_N)Lu|| + ||P_N L(I-P_N)u|| .$$

The first term on the right goes to zero as $N \to \infty$ at a rate governed solely by the smoothness of Lu; it measures the error in the N term Legendre-Galerkin expansion of Lu. The second term is estimated as follows. Set

$$L(I-P_N)u = \sum_{n=1}^{\infty} a_n \phi_n(x)$$

where $\{\phi_n\}$ are normalized Legendre polynomials. If L is a finite-order differential operator so L^* is also a finite-order differential operator (for example, $L^*=\partial/\partial x$ if $L=-\partial/\partial x$), then

$$a_n = (\phi_n, L(I-P_N)u)$$

$$= (L^*\phi_n, (I-P_N)u) .$$

Thus,

$$|a_n| \le ||L^*\phi_n|| \, ||(I-P_N)u||$$
$$= O(n^A/N^B) \quad (n\to\infty ; N\to\infty),$$

where A depends only on L ($A = 3/2$ if $L = -\partial/\partial x$ and ϕ_n is a normalized Legendre polynomial) and B depends only on the smoothness of u (B is arbitrary if u is infinitely differentiable). Thus,

$$\|P_N L(I-P_N)u\|^2 = \sum_{n=1}^{N} a_n^2 \to 0$$

faster than any power of $1/N$ as $N\to\infty$ if u and all its derivatives are smooth. This proves consistency. This kind of proof extends to a wide variety of the examples to be discussed in Sects. 7 and 8, but will not be repeated.

(ii) Solution of $u_t = xu_x$ with the boundary conditions $u(\pm 1,t) = 0$ is a well posed problem in the Chebyshev inner product

$$(u,v) = \int_{-1}^{1} \frac{u(x)v(x)}{(1-x^2)^{\frac{1}{2}}}\,dx .$$

In fact, if $L = x\,\partial/\partial x$, u is differentiable, and $u(\pm 1) = 0$ then, by integration by parts,

$$(u,Lu) = \int_{-1}^{1} x(1-x^2)^{-\frac{1}{2}}\,u\,u_x dx = -\int_{-1}^{1} (1-x^2)^{-\frac{3}{2}} u^2\,dx \leq 0 .$$

Thus, Galerkin approximation to the problem is stable using Chebyshev polynomials.

(iii) Solution of $u_t + u_x = 0$ $(0 \leq x < \infty)$ with $u(0,t) = 0$ is a well posed problem in the Laguerre inner product

$$(u,v) = \int_0^\infty u(x)v(x)e^{-x}dx \ .$$

In fact, if $u(0,t) = 0$ then, by integrating by parts,

$$-\int_0^\infty uu_x e^{-x}dx = -\frac{1}{2}e^{-x}u^2 \Big|_0^\infty - \int_0^\infty e^{-x}u^2dx < 0 \ .$$

Similarly, the problem $u_t = u_{xx}$ $(0 \leq x < \infty)$ with $u(0,t) = 0$ is also stable in the Laguerre norm.

(iv) Solution of $u_t = -xu_x$ $(-\infty < x < \infty)$ is well posed in the Hermite inner product

$$(u,v) = \int_{-\infty}^\infty u(x)v(x)e^{-x^2}dx \ .$$

In fact,

$$\frac{\partial}{\partial t}(u,u) = -2\int_{-\infty}^\infty x\, e^{-x^2}\, uu_x\, dx,$$

so that integration by parts gives

$$\frac{\partial}{\partial t}(u,u) = \int_{-\infty}^\infty u^2e^{-x^2}(1-2x^2)dx \leq (u,u),$$

where we assume that $u << x^{-\frac{1}{2}} \exp\left(\frac{1}{2}x^2\right)$ as $|x| \to \infty$.

(v) The heat equation $u_t = u_{xx}$ with $u(\pm 1,t) = 0$ is semi-bounded in the Chebyshev norm. In fact, if u is differentiable for $|x| \leq 1$ then

$$\int_{-1}^{1} (1-x^2)^{-\frac{1}{2}} uu_{xx}\, dx = (1-x^2)^{-\frac{1}{2}} uu_x \Big|_{-1}^{1} - \int_{-1}^{1} [u(1-x^2)^{-\frac{1}{2}}]_x u_x\, dx$$

The first term vanishes because u is a polynomial in x and therefore $u(\pm 1) = 0$ implies

$$\frac{u}{(1-x^2)^{1/2}} \Bigg|_{x=\pm 1} = 0$$

The integral term on the right is

$$- \int_{-1}^{1} [u(1-x^2)^{-\frac{1}{2}}]_x\, u_x\, dx$$

$$= - \int_{-1}^{1} [u(1-x^2)^{-\frac{1}{2}}]_x [u(1-x^2)^{-\frac{1}{2}}]_x (1-x^2)^{\frac{1}{2}}\, dx$$

$$+ \frac{1}{2} \int_{-1}^{1} \frac{\partial}{\partial x} \left[u(1-x^2)^{-\frac{1}{2}} \right]^2 x(1-x^2)^{-\frac{1}{2}}\, dx$$

$$= - \int_{-1}^{1} \left[(u(1-x^2)^{-\frac{1}{2}})_x \right]^2 (1-x^2)^{\frac{1}{2}}\, dx$$

$$+ \frac{1}{2} u^2 x(1-x^2)^{-\frac{3}{2}} \Big|_{-1}^{1} - \frac{1}{2} \int_{-1}^{1} u^2 (1-x^2)^{-\frac{5}{2}}\, dx \tag{7.3}$$

$$\leq 0$$

and therefore

$$\frac{d}{dt} \int_{-1}^{1} \frac{u^2}{\sqrt{1-x^2}}\, dx \leq 0 .$$

In the next three examples we generalize the proofs of stability and convergence for Galerkin approximations given in Example 7.1 to show the stability and convergence of tau approximations.

Example 7.2: <u>A semi-bounded tau approximation</u>

Consider the equation

$$\frac{\partial u}{\partial t} = x \frac{\partial u}{\partial x}$$

with the boundary conditions

$$u(\pm 1,t) = 0.$$

It was shown in Example 7.1(ii) that if $L = x\partial/\partial x$, then

$$L + L^* \leq 0$$

in the Chebyshev inner product. If we seek the solution as the truncated Chebyshev series

$$u_N = \sum_{n=0}^{N} a_n T_n$$

by the tau method, then u_N satisfies exactly the equation

$$\frac{\partial u_N}{\partial t} - x \frac{\partial u_N}{\partial x} = \tau_N(x)T_N(x) + \tau_{N-1}(t)T_{N-1}(x) \qquad (7.4)$$

Equating coefficients of x^N and x^{N-1} on both sides of (7.4), we obtain

$$a_N' - Na_N = \tau_N$$

$$a_{N-1}' - (N-1)a_{N-1} = \tau_{N-1}$$

since $T_n = 2^{n-1}x^n - n2^{n-3}x^{n-2} + \cdots$. Therefore,

$$\frac{2}{\pi}\left(u_N, \frac{\partial u_N}{\partial t}\right) = \frac{1}{\pi}([L+L^*]u_N, u_N) + [a_N' - Na_N]a_N$$
$$+ [a_{N-1}' - (N-1)a_{N-1}]\, a_{N-1} \tag{7.5}$$

so that

$$\tfrac{1}{2}\frac{\partial}{\partial t}\left[\frac{2}{\pi}(u_N, u_N) - a_N^2 - a_{N-1}^2\right] = \frac{1}{\pi}([L+L^*]u_N, u_N) - Na_N^2 - (N-1)a_{N-1}^2 \leq 0.$$

Since

$$(u_N, u_N) = \frac{\pi}{2}\sum_{n=0}^{N} c_n a_n^2 ,$$

with $c_0=2$, $c_n=1\,(n\geq 1)$, the above inequality is equivalent to

$$\frac{\partial}{\partial t}\sum_{n=0}^{N-2} c_n a_n^2 \leq 0 \tag{7.6}$$

This proves stability: a_N and a_{N-1} are bounded because they are determined in terms of $a_0, a_1, \ldots, a_{N-2}$ by the boundary conditions $u(\pm 1, t) = 0$.

For this example, we can prove stability directly from the matrix representation of L_N. In fact,

$$(L_N)_{jk} = \frac{1}{c_j}[-k\delta_{jk} + 2\sum_{\substack{\ell=0 \\ \ell \text{ even}}}^{N-j} k\,\delta_{j+\ell,k}] \quad (0 \le j \le N,\ 0 \le k \le N), \tag{7.7a}$$

In the tau approximation, the boundary conditions $u(\pm 1,t) = 0$ require that the last two rows of the matrix L_N be replaced by

$$(L_N)_{N-1,k} = (-1)^k, \tag{7.7b}$$

$$(L_N)_{N,k} = 1. \tag{7.7c}$$

If the boundary conditions (7.7b,c) are not applied then the spectral approximation is unstable: without the boundary conditions L_N has the eigenvalue N [with the eigenvector $a_{N-2k} = \binom{N}{k}$, $a_{N-2k-1} = 0$] so that

$$\|e^{L_N t}\| \ge e^{Nt} .$$

To prove convergence when the boundary conditions (7.7b,c) are applied, let us first consider an odd solution in which $a_n = 0$ if n is even. If we assume that $N = 2M+1$ and set

$$d_k = a_{2k+1} \qquad (0 \le k \le M)$$

then the system reduces to

$$\frac{\partial \vec{d}}{\partial t} = D\,\vec{d}$$

where

$$D_{jk} = -(2k+1)\delta_{jk} + 2\sum_{\ell=0}^{M-j} (2k+1)\delta_{j+\ell,k} - 2N \quad (0 \le j < M,\ 0 \le k < M)$$

If we introduce the M x M transformation matrix S defined by

$$S_{jk} = \delta_{jk} - \delta_{j+1,k} \quad (0 \le j < M,\ 0 \le k < M\ )$$

then $S(D+D^*)S^*$ is a diagonal matrix with entries $(-4, -4, \ldots, -4, -4N - 12)$. Thus, we obtain $D + D^* \leq 0$, so that $\partial(\vec{d},\vec{d})/\partial t \leq 0$ which proves stability.

Example 7.3. Stability of tau methods applied to degree-reducing semi-bounded equations

An argument similar to that given in Example 7.2 demonstrates stability of tau methods in terms of arbitrary orthonormal polynomial bases for equations $\frac{\partial u}{\partial t} = Lu$ where L is semi-bounded and degree reducing: L is said to be degree reducing if for any polynomial P_N of degree N, LP_N is a polynomial of degree at most $N - k$ where k is the number of boundary conditions that are applied. If L is degree reducing, equating coefficients of $x^{N-k+2}, \ldots, x^N$ in

$$\frac{\partial u_N}{\partial t} = L\, u_N + \sum_{n=N-k+1}^{N} \tau_n \phi_n$$

implies that $\tau_n(t) = a_n'(t)$ for $n = N-k+1, \ldots, N$; here

$$u_N(x,t) = \sum_{n=0}^{N} a_n(t)\phi_n(x)$$

and the orthonormal expansion polynomial $\phi_n(x)$ is assumed of degree n. Therefore,

$$\tfrac{1}{2}\frac{\partial}{\partial t}(u_N, u_N) - \sum_{n=N-k}^{N} a_n' a_n = ([L+L^*]\, u_N, u_N) \leq 0$$

so that

$$\frac{\partial}{\partial t} \sum_{n=0}^{N-k} a_n^2 \leq 0.$$

which proves stability since $a_{N-k+1},\dots,a_N$ are determined by the boundary conditions in terms of $a_0,a_1,\dots,a_{N-k}$.

Example 7.3: More stable tau approximations

(i) Suppose that

$$u_t + u_x = 0 \quad (-1 \le x \le 1,\ t > 0)$$

$$u(-1,t) = 0$$

is solved by tau approximation using Legendre polynomials. The Nth degree Legendre-tau approximation u_N satisfies

$$\frac{\partial}{\partial t} u_N + \frac{\partial}{\partial x} u_N = a_N' P_N \quad , \quad u_N(-1,t) = 0.$$

Since $\int_{-1}^{1} P_N^2(x)\,dx = 2/(2N+1)$ [see A.25], we obtain

$$\frac{d}{dt}\left[\int_{-1}^{1} u_N^2\,dx - \frac{2}{2N+1}\, a_N^2\right] \le 0 ,$$

which proves stability.

(ii) Suppose that

$$u_t = u_{xx}$$

$$u(\pm 1,t) = 0$$

is solved by the tau method using Chebyshev polynomials. Since $L = \frac{\partial^2}{\partial x^2}$ is degree-reducing and $L + L^* \le 0$ [see Example 7.1(v)], the method is stable.

(iii) The solution of

$$u_t + u_x = 0 \qquad (0 \le x \le \infty\ ,\ t > 0)$$

$$u(0,t) = \sin t \qquad (t > 0) \tag{7.8}$$

$$u(x,0) = 0 \qquad (0 \le x \le \infty)$$

by Laguerre polynomials is stable using the tau method since, by Example 7.1 (iii), L is semi-bounded. The equations of the Laguerre-tau approximation to (7.8) are a simple modification of (2.23-24). In Fig. 7.1 we compare this tau approximation with the exact solution of (7.8) at $t = 30$ for a 20-term Laguerre expansion. The reader should compare this approximate result obtained by the tau method with the best Laguerre approximation to $\sin x$ plotted in Fig. 3.12.

In the next example we discuss some ways to find non-trivial Liapounov matrices $\{H_N\}$ when L is not semi-bounded.

Example 7.5: <u>Polynomial approximations to a variable coefficient hyperbolic equation</u>

Consider the initial-value problem

$$u_t = -xu_x \qquad |x| < 1$$

$$u(x,0) = g(x) \tag{7.9}$$

which is well posed without requiring any boundary conditions. The exact solution to this problem is

$$u(x,t) = g(xe^{-t})$$

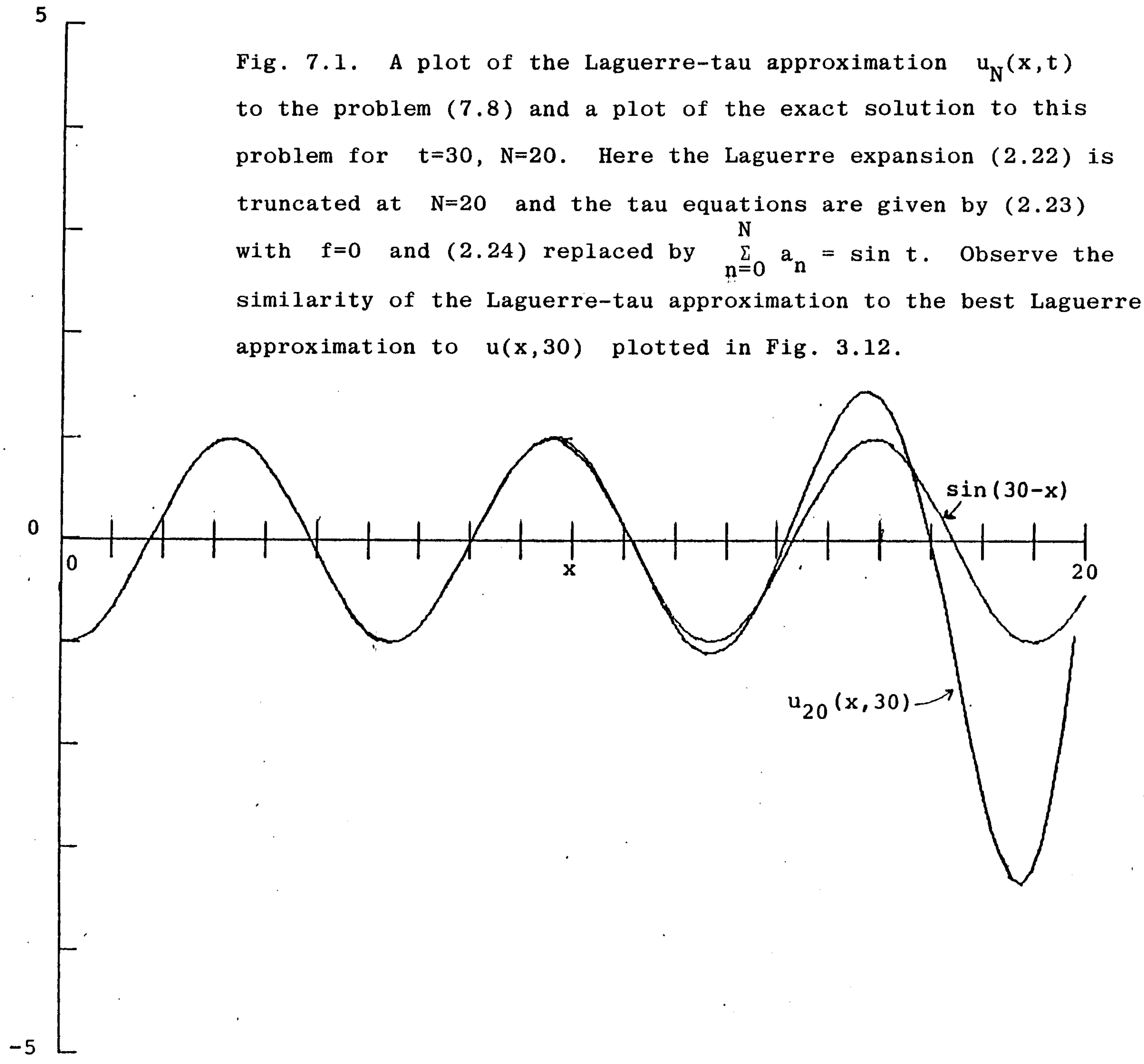


Fig. 7.1. A plot of the Laguerre-tau approximation $u_N(x,t)$ to the problem (7.8) and a plot of the exact solution to this problem for $t=30$, $N=20$. Here the Laguerre expansion (2.22) is truncated at $N=20$ and the tau equations are given by (2.23) with $f=0$ and (2.24) replaced by $\sum_{n=0}^{N} a_n = \sin t$. Observe the similarity of the Laguerre-tau approximation to the best Laguerre approximation to $u(x,30)$ plotted in Fig. 3.12.

so that $u(x,t)$ approaches a constant as $t\to\infty$:

$$u(x,t) \sim g(0) \quad (t\to\infty,\ |x|\leq 1).$$

The problem is well-posed in the sense that $||\exp(Lt)||$ is finite for finite t, where $L= - x\partial/\partial x$ and $||\cdot||$ is the usual L_2 norm. However, $||\exp\ (Lt)||= \exp\ (\frac{1}{2}\ t)$ so $||\exp(Lt)||$ is unbounded as $t\to\infty$. To show this, we observe that the function that extremizes $||u(t)||$ subject to $||u(0)||= 1$ satisfies $u(x,0) = g_t(x)$ where

$$g_t(x) = \begin{cases} \pm \dfrac{e^{t/2}}{\sqrt{2}} & |x|\leq e^{-t} \\ 0 & |x|>e^{-t}. \end{cases}$$

The operator L is semibounded in the usual L_2 norm:

$$\begin{aligned} \frac{\partial}{\partial t} \int_{-1}^{1} u^2\, dx &= - \int_{-1}^{1} x\, \frac{\partial u^2}{\partial x}\, dx = - u^2(1) - u^2(-1) + \int_{-1}^{1} u^2\, dx \\ &\leq \int_{-1}^{1} u^2\, dx, \end{aligned}$$

so $L + L^* \leq I$. Therefore, Galerkin polynomial solution of (7.9) is stable and convergent. The Legendre polynomial approximation $u_N(x,t)$ satisfies

$$\frac{\partial u_N}{\partial t} + x\, \frac{\partial u_N}{\partial x} = 0 \tag{7.10}$$

exactly because no boundary conditions are applied and L is degree preserving. Therefore, Galerkin, tau, and collocation approximations to (7.9) are identical and all three methods are stable.

In fact, all polynomial-spectral methods applied to (7.9) satisfy (7.10); all polynomial methods for this problem give identical results and, therefore, they are all stable in the usual L_2 norm. In terms of the natural norms for a general polynomial basis $\{\psi_n\}$, i.e. that norm in which $(\psi_i, \psi_j) = \delta_{ij}$, the spectral approximation (7.10) is algebraically stable if the $N \times N$ matrix whose elements are

$$(H_N)_{jk} = \int \psi_j(x) \; \psi_k(x) \; dx$$

has a condition number which is bounded algebraically, i.e., $\| H_N \| \, \| H_N^{-1} \| = O(N^{\beta}) \quad (N \to \infty)$.

As an example of the complicated behavior of spectral approximations for this problem in norms different from the usual L_2 norm, let us consider the Chebyshev-L_2 norm. It may easily be shown that $L + L^*$ is not semibounded in the Chebyshev inner product. For example, consider the trial function

$$v = T_0 - T_{2N}$$

satisfies

$$
\begin{aligned}
([L+L^*]v,v) &= -(xv_x,v) - (v,xv_x) \\
&= -\left(-2N[T_{2N-1}+\dots+T_1],T_1 - \frac{T_{2N+1}+T_{2N-1}}{2}\right) \\
&= \frac{1}{3} N(v,v) .
\end{aligned}
$$

Nevertheless, Chebyshev approximation to this problem is algebraically stable. We will demonstrate this fact explicitly by construction of a Liapounov matrix.

A Liapounov matrix for the Chebyshev approximation to (7.9) may be found by direct examination of the evolution equation for the vector $\vec{a}_N = (a_0,\dots,a_N)$:

$$
\frac{\partial a_n}{\partial t} = -n\, a_n - 2 \sum_{\substack{p=n+2 \\ p+n \text{ even}}}^{N} p a_p \qquad (n = 0,\dots,N). \qquad (7.11)
$$

Since a_0 decouples from $a_1,\dots,a_N$ in (7.11), we can restrict attention to $a_1,\dots,a_N$. Suppose we define the matrix H_N by

$$
(H_N)_{jk} = \frac{1}{j}\delta_{jk} , \qquad (1 \le j,k \le N).
$$

H_N is the Liapounov matrix for L_N^T: in fact, (5.7b) is satisfied because

$$\frac{1}{2}(H_N L_N^T + L_N H_N) = \begin{pmatrix} -1 & 0 & -1 & 0 & -1 & \cdots \\ 0 & -1 & 0 & -1 & 0 & \cdots \\ -1 & 0 & -1 & 0 & -1 & \cdots \\ 0 & -1 & 0 & -1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \le 0 .$$

The matrix displayed above has rank 2 and the nonzero eigenvalues are $-[N/2]$, $-[(N+1)/2]$. Therefore, by the theory of Sec. 5,

$$\| e^{L_N t} \| \le \sqrt{\|H_N\| \, \|H_N^{-1}\|} = \sqrt{N} \quad ,$$

where $\|\cdot\|$ is now the Chebyshev norm. Thus, L_N is algebraically stable in the Chebyshev norm even though $L_N + L_N^*$ is unbounded in this norm.

The qualitative behavior of the Chebyshev norm of $\exp(L_N t)$ as a function of N and t is as follows. For fixed t and $N \to \infty$, $\|\exp(L_N t)\| = 0\,(N^{1/4})$; this result is justified heuristically by following the argument given in Sec. 5 that led to (5.4). On the other hand if $t \gtrsim \ln N$, $\|\exp(L_N t)\| = 0(N^{1/2})$ as $N \to \infty$. A heuristic justification of this result is as follows. Let $u(x,0) = 1$ for $|x| \le \varepsilon$, 0 for $|x| > \varepsilon$. Then the exact solution of (7.9) for $t > \ln 1/\varepsilon$ is $u(x,t) \sim 1$ for $|x| \le 1$, so $\|u(x,t)\|^2 \sim \pi$ as $\varepsilon \to 0^+$ for $t > \ln 1/\varepsilon$. As in Sec. 5, we conclude that $\|\exp(L_N t)\| = 0\,(N^{1/2})$ for $t \gtrsim \ln N$ as $N \to \infty$. [Even in the usual L_2 norm, $\|\exp(L_N t)\| = 0\,(N^{1/2})$ when $t \gtrsim \ln N$, which mimics the unbounded growth of $\|\exp(Lt)\|$ as $t \to \infty$.]

8. Constant Coefficient Hyperbolic Equations

In this Section, we discuss the stability of spectral methods for the problem

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0 \qquad (|x| \leq 1, \quad t > 0) \qquad (8.1)$$

with the initial condition

$$u(x,0) = f(x) \qquad (|x| \leq 1) \qquad (8.2)$$

and the boundary condition

$$u(-1,t) = 0 \qquad (t > 0) \quad . \qquad (8.3)$$

The results for this problem can be extended to a general hyperbolic system of the form

$$u_t = Au_x$$

with characteristic boundary conditions, because for any hyperbolic system A can be diagonalized by a real similarity tranformation.

The operator $L = -\frac{\partial}{\partial x}$ is semi-bounded in the usual $L_2(-1,1)$ norm when operating on the subspace of functions v that satisfy the boundary condition $v(-1,t) = 0$. In fact

$$(v,[L+L^*]v) = -2\int_{-1}^{1} v \frac{\partial v}{\partial x} dx = -v^2(1) \leq 0$$

and therefore Galerkin and tau methods are stable using Legendre polynomials.

However, L is not semi-bounded in the Chebyshev norm. To show this, we set

$$v(x) = T_{2N}(x) - T_1(x) - 2T_0(x)$$

so that $v(-1) = 0$. Since

$$T'_{2N} = 2N[T_{2N-1}+T_{2N-3}+\cdots+T_1] ,$$

we find

$$\begin{aligned}(v,[L+L^*]v) &= -2\int_{-1}^{1}(1-x^2)^{-\frac{1}{2}}\frac{\partial v}{\partial x}\, v\, dx \\ &= -2\int_{-1}^{1}(1-x^2)^{-\frac{1}{2}}[2N(T_{2N-1}+T_{2N-3}+\ldots+T_1)-T_0](T_{2N}-T_1-2T_0)\,dx \\ &= \frac{2N-4}{5}\;(v,v). \end{aligned} \tag{8.4}$$

The fact that $L+L^*$ is not semi-bounded is consistent with the fact that $\exp(Lt)$ is not a bounded operator for $t<2$ in the Chebyshev norm (see Sec. 5). However, these results do not prove that Chebyshev-spectral approximation to (8.1-3) is not convergent. In fact, we shall show that, while Chebyshev-spectral approximation to (8.1-3) is not stable in the Chebyshev L_2 norm, it is algebraically stable in this norm.

In order to investigate algebraic stability, we must study more carefully the behavior of the Chebyshev coefficients of the approximate solution

$$u_N = \sum_{n=0}^{N} a_n(t) T_n(x) \quad .$$

The differential equations for the a_n's are given by (2.11) for Galerkin approximation, (2.19) for the tau method, and (2.32) for the collocation method. As remarked in Sec. 2, all these equations may be written in the vector form

$$\frac{\partial \vec{a}}{\partial t} = L_N \vec{a}$$

where $\vec{a} = (a_0, a_1, \ldots a_{N-1})$ and L_N is an $N \times N$ matrix. The value of $a_N(t)$ is determined in terms of $\vec{a}(t)$ by the boundary condition (8.3).

Numerical Evidence for Algebraic Stability

Let us first examine the behavior of $L_N + L_N^*$. In Table 8.1 we list the largest eigenvalue of $L_N + L_N^*$ for $N = 10, 20, \ldots, 100$ for the three Chebyshev methods. This table indicates that the largest positive eigenvalue of $L_N + L_N^*$ grows like cN^2 for some constant c. If L_N were a normal matrix this would imply that $\|e^{L_N t}\|$ behaves like $\exp(\frac{1}{2} c N^2 t)$. However, the matrices L_N are not normal and therefore the large eigenvalues of $L_N + L_N^*$ do not imply instability.

Table 8.1

N	Collocation	Tau	Galerkin
10	68.8413	21.4089	72.8947
20	287.6920	84.8970	296.3027
30	656.4818	190.4908	669.6434
40	1175.2124	338.1769	1192.9231
50	1843.8839	527.9525	1866.1433
60	2662.4966	759.8167	2689.3042
70	3631.0503	1033.7690	3662.4061
80	4749.5453	1349.8093	4785.4489
90	6017.9812	1707.9375	6058.4329
100	7436.3584	2108.1534	7481.3579

Table 8.1. The largest positive eigenvalue λ_{max} of $L_N + L_N^*$ for the Chebyshev-spectral solution of the one-dimensional wave equation (8.1-3). The Galerkin approximation to this problem is given by the solution to (2.11), the tau approximation is given by (2.19), and the collocation approximation is given by (2.32). Observe that $\lambda_{max} \sim cN^2$ as $N \to \infty$ where $c \doteqdot 0.75$ for the Galerkin and collocation methods and $c \doteqdot 0.21$ for the tau method.

In Table 8.2, we give the norms of the matrices $\exp[L_N] \cdot \exp[L_N^*]$ for the three projection methods (Galerkin, collocation, and tau). The results indicate that $||\exp(L_N)||$ grows only like $N^{1/4}$ as $N \to \infty$ (as argued heuristically in Sec.5). In other words, L_N is algebraically stable (at least for t=1). This result shows the extreme pessimism of the energy estimate $||\exp(L_N)|| = 0(\exp(\frac{1}{2} c N^2))$; crude energy methods may be very misleading for non-normal evolution operators.

In order to understand better how the Chebyshev spectral methods avoid an energy 'catastrophe' [energy growth like $\exp(cN^2 t)$] we have solved the tau equations (2.19) numerically with a very 'bad' initial condition:

$$u_N(x,0) = [T_N(x) + 2T_{N-1}(x) + (-1)^N T_0(x)]/\sqrt{7} . \qquad (8.5)$$

For the tau method, this initial condition satisfies

$$\frac{\partial}{\partial t} (u_N, u_N) \Big|_{t=0} = (u_N, (L_N + L_N^*) u_N) = 0(N^2) \qquad (N \to \infty) .$$

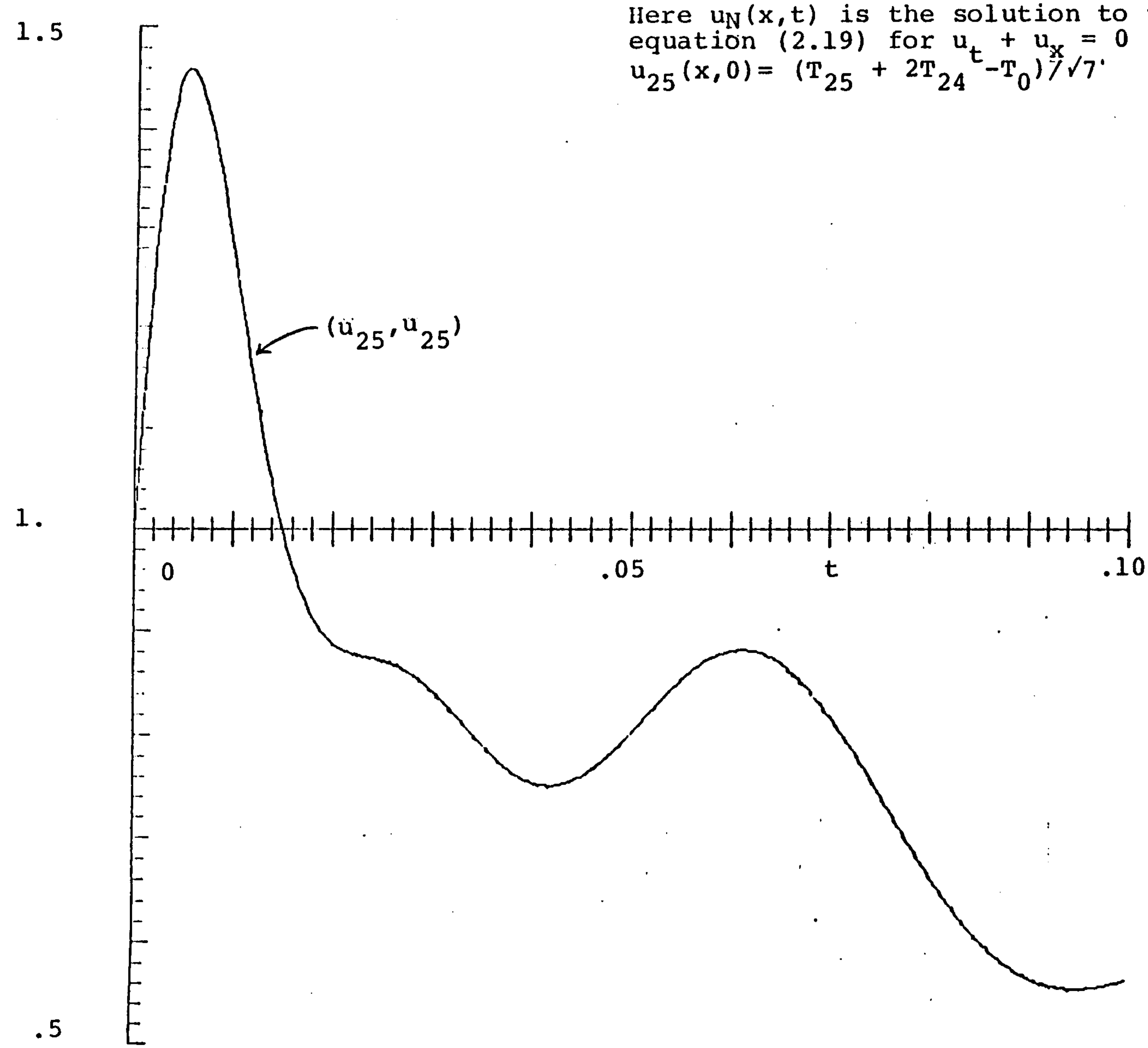
In Figs. 8.1-2 we plot the energy (u_N, u_N) vs t for $N = 25$ and $N = 50$. It is apparent that the initial slope of the energy growth is of order N^2 but that the energy does not maintain this rapid rate of growth. Observe that the region of rapid growth is closer to $t = 0$ for $N = 50$ than for $N = 25$. The behavior observed in Figs. 8.1-2 is not inconsistent with the fact that $u_N(t = 0)$ is a 'bad' eigenmode of $L_N + L_N^*$. Because L_N is

Table 8.2

N	Collocation	Tau	Galerkin
10	2.0707	2.0003	2.5788
20	2.7932	2.8119	3.1903
30	3.4620	3.4857	3.8328
40	4.0324	4.0514	4.4078
50	4.5222	4.5339	4.8630
60	4.9117	4.9855	5.2057
70	5.2961	5.4002	5.5262
80	5.6586	5.7770	5.8689
90	6.0282	6.1401	6.2526
100	6.3818	6.4831	6.6257

Table 8.2. The largest eigenvalue λ_{max} of $\exp(L_N)\exp(L_N^*)$. Observe that λ_{max} behaves as $cN^{1/2}$ as $N \to \infty$ where $c \doteqdot 0.6$ for all three spectral methods. The largest eigenvalue of $\exp(L_N)\ \exp(L_N^*)$ grows only like $N^{1/2}$ despite the existence of eigenvalues of $L_N + L_N^*$ growing like N^2 (see Table 8.1).

Fig. 8.1. A plot of the energy (u_N, u_N) vs t for the 'bad' initial conditions (8.5) with $N = 25$. Here $u_N(x,t)$ is the solution to the Chebyshev-tau equation (2.19) for $u_t + u_x = 0$ with $N=25$, $u_{25}(x,0) = (T_{25} + 2T_{24} - T_0)/\sqrt{7}$.



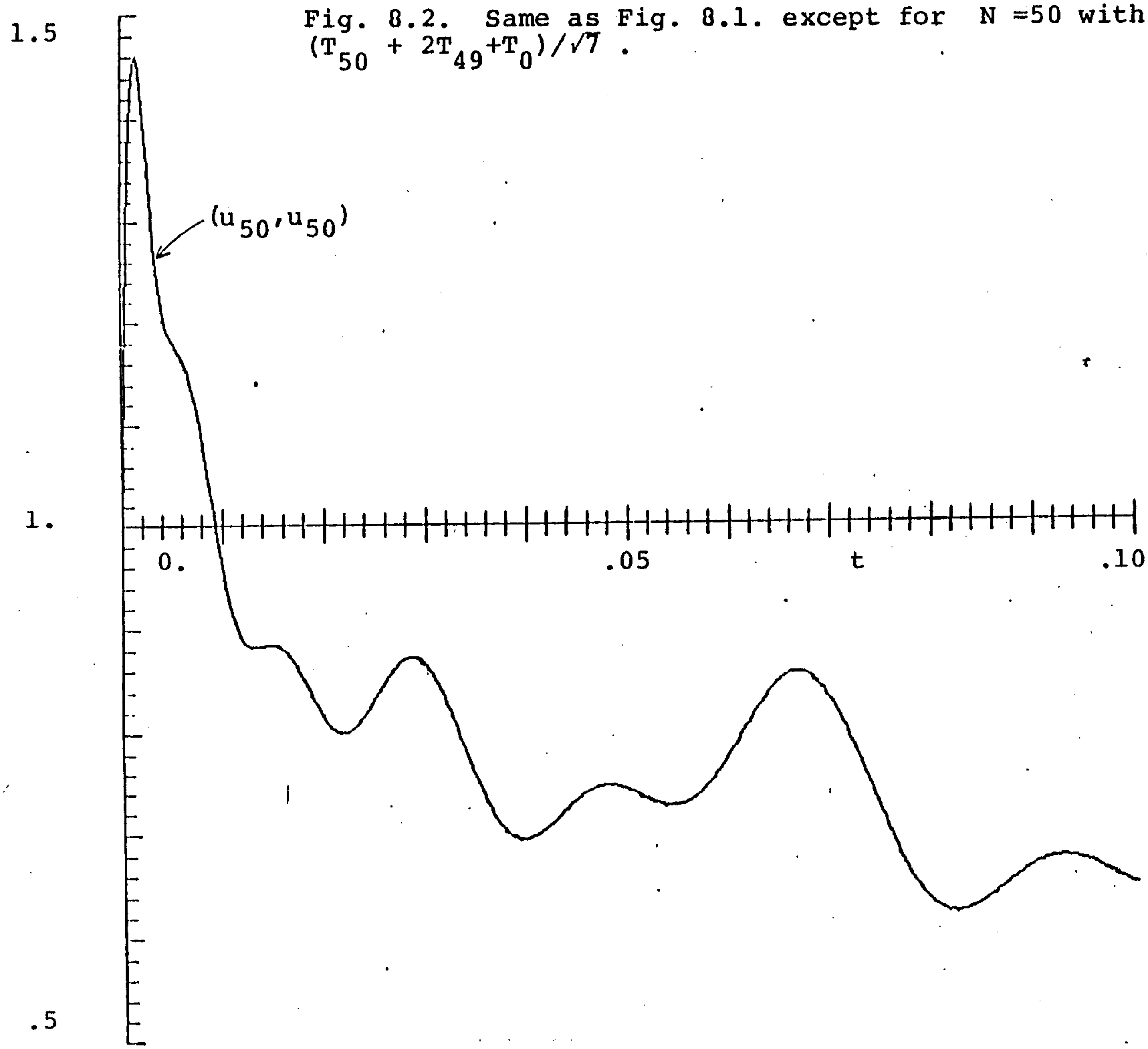


Fig. 8.2. Same as Fig. 8.1. except for $N = 50$ with $u_{50}(x,0) = (T_{50} + 2T_{49} + T_0)/\sqrt{7}$.

non-normal the 'bad' initial condition is not an eigenmode of L_N so that after evolution from 0 to t $\exp(L_N t)$, u_N 'rotates' out of the region of bad modes of $L_N + L_N^*$.

The direct computation of $\exp[L_N t]$ for $t=1$ is not enough to verify algebraic stability because the theory of Sec. 5 shows that we must study the behavior of $\exp[L_N t]$ for a complete time interval $0 \leq t \leq T$. This may be done using the method suggested in Sec. 5 for the numerical verification of algebraic stability. First, in Table 8.3 we list the numerically computed eigenvalues of L_N. Observe that all the eigenvalues of L_N have negative real part. (This result will be shown rigorously later.) Therefore, $||\exp(L_N t)|| \to 0$ as $t \to \infty$ for fixed N. Thus the Chebyshev approximations are asymptotically stable in the sense that they remain bounded as $t \to \infty$ with N fixed.

In Figs. 8.3-5, we plot the L_1-matrix norm of $\exp(L_N t)$ vs t for N=5,15,25. Observe that as $t \to \infty$ for fixed N, $||\exp(L_N t)||_1$ approaches zero while it grows slowly (like $N^{1/2}$) as $N \to \infty$. for fixed $t<2$ (Note that growth of $||\exp(L_N t)||_1$ like $N^{1/2}$ as $N \to \infty$ is not inconsistent with growth of $||\exp(L_N t)||_2$ like $N^{1/4}$.) Also observe that the norms seem to have a boundary layer at $t=2$ such that $||\exp(L_N t)||_1 \to \infty$ as $N \to \infty$ for $t<2$ and $\to 0$ as $N \to \infty$ for $t>2$. This behavior is consistent with the unboundedness of $\exp(Lt)$ for $t<2$ [see (5.4)].

Asymptotic stability does not prove <u>stability</u> because L_N is not normal. The next step in the computational proof of stability is to compute numerically the Liapounov matrices H_N satisfying

Table 8.3

N	Galerkin	Tau	Collocation
10	-2.4532	-2.9994	-1.9306
20	-2.5932	-3.9320	-2.1591
30	-2.7267	-4.5380	-2.3247
40	-2.8495	-4.9918	-2.4659
50	-2.9669	-5.3837	-2.5965
60	-3.0824	-5.7266	-2.7226
70	-3.1985	-6.0489	-2.8478
80	-3.3162	-6.3650	-2.9738
90	-3.4365	-6.6861	-3.1017
100	-3.5597	-7.0229	-3.4335

Table 8.3. The real part of the eigenvalue of L_N with least negative real part for the collocation, tau, and Galerkin spectral approximations to (8.1.3). Since all the eigenvalues of L_N have negative real parts, these spectral methods are asymptotically stable as $t \to \infty$.

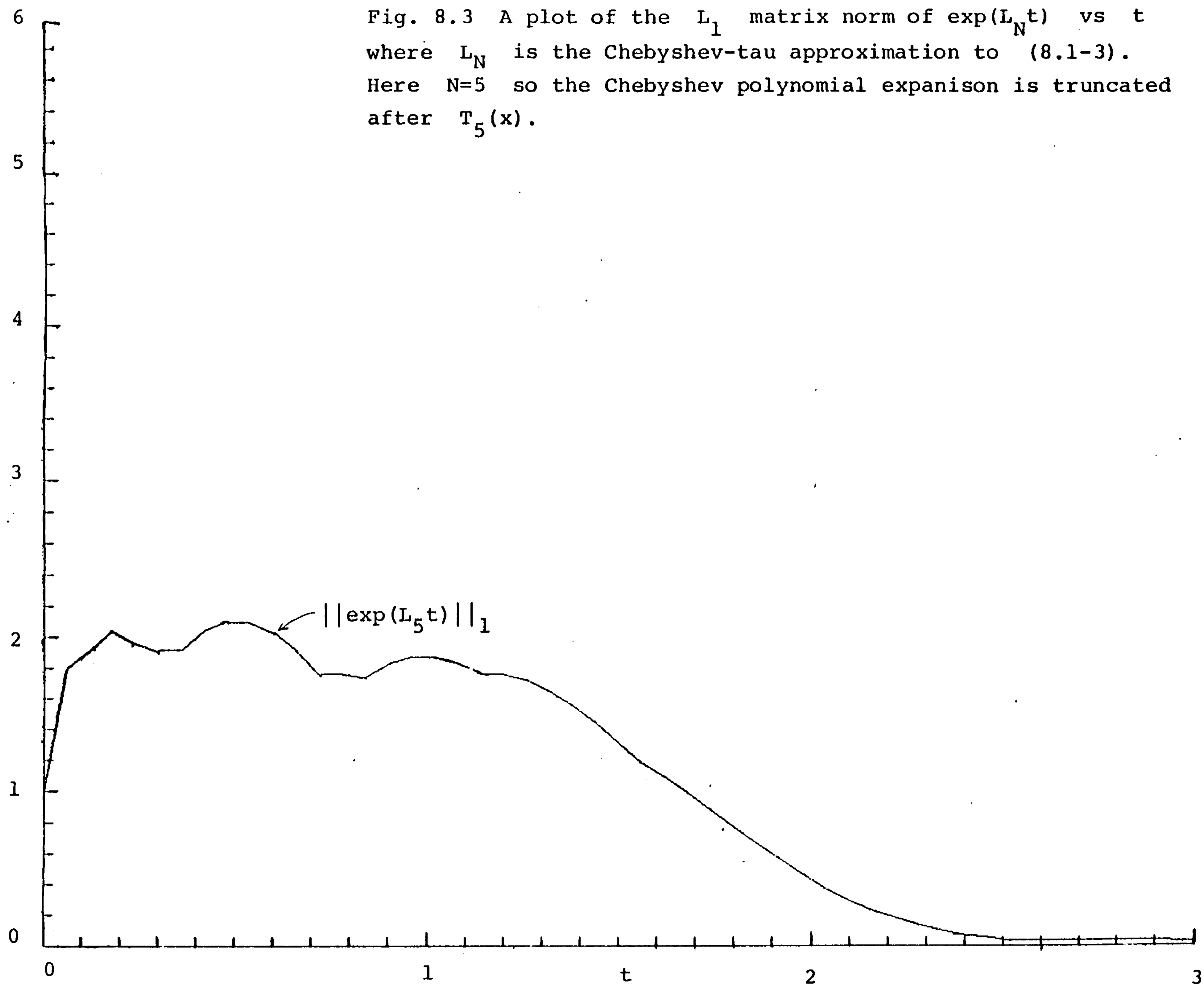
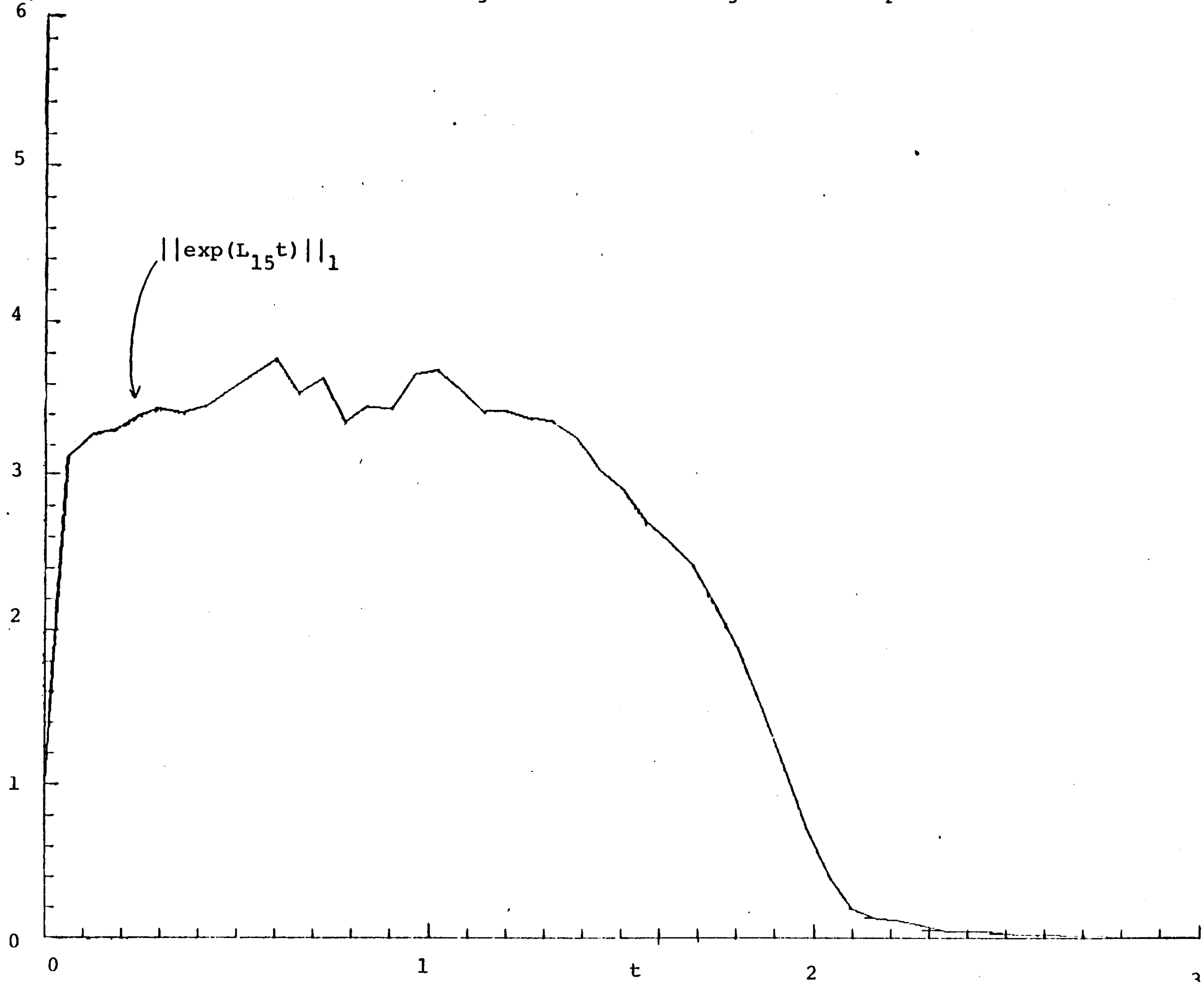


Fig. 8.3 A plot of the L_1 matrix norm of $\exp(L_N t)$ vs t where L_N is the Chebyshev-tau approximation to (8.1-3). Here $N=5$ so the Chebyshev polynomial expanison is truncated after $T_5(x)$.

Fig. 8.4. Same as Fig. 8.3 except N=15.



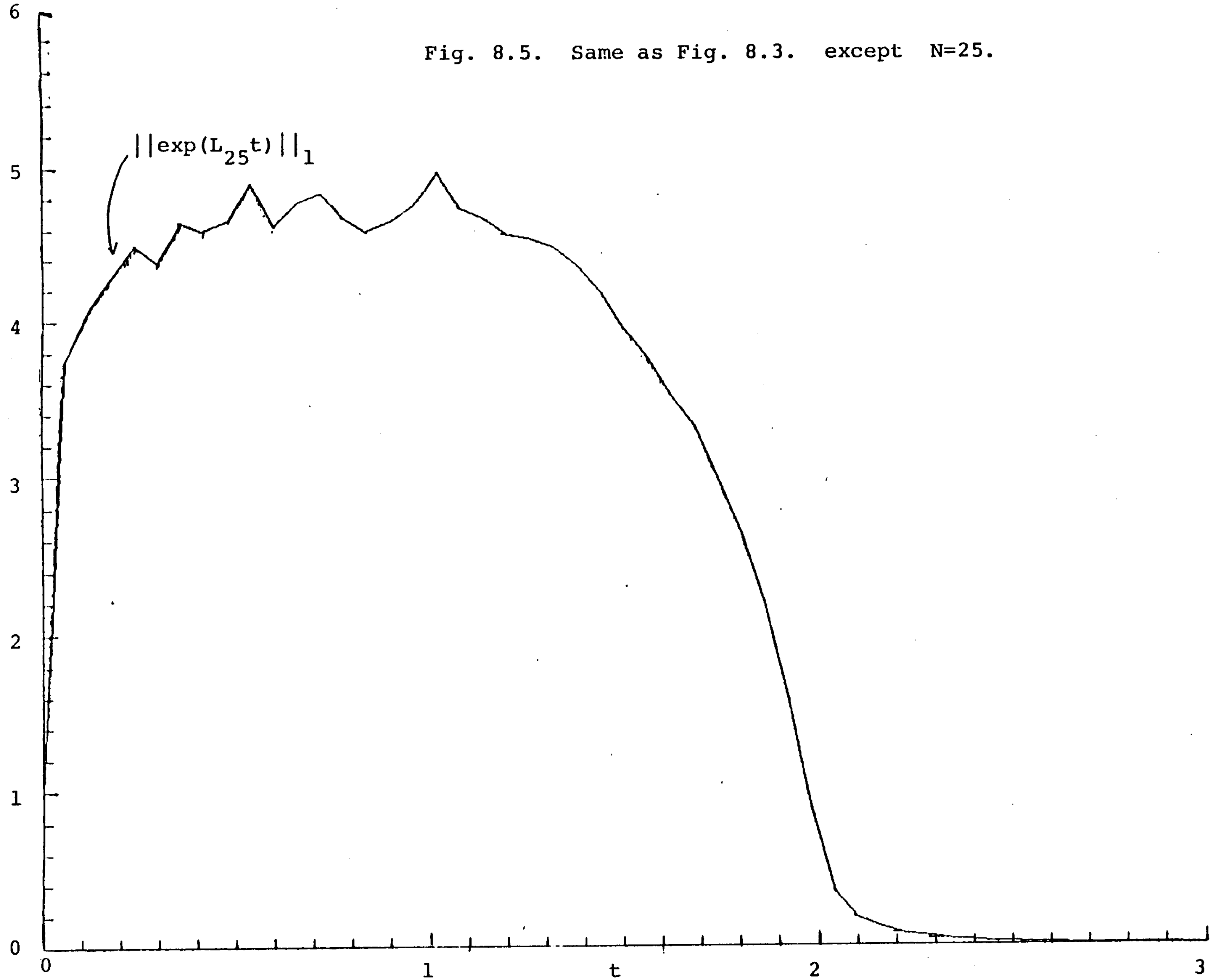


Fig. 8.5. Same as Fig. 8.3. except N=25.

$$H_N A_N + A_N^* H_N = - I \qquad (8.6)$$

A good method to compute H_N is described by Bartels & Stewart (1974). In Table 8.4 we list the condition number of H_N for the Galerkin, collocation and tau methods. This table suggests that the condition number of H_N grows at most like N^3 as N for the Galerkin and collocation methods[†] and like N^2 for the tau method. Recalling (5.11), we obtain

$$||[\exp[L_N t]|| = 0\ (N^{\frac{3}{2}} e^{-\frac{t}{2}}) \ . \qquad (8.7)$$

for all three methods. It should be noted that (8.7) gives only an upper bound for $||\exp[L_N t]||$. According to the theory given in Sec. 5, this upper bound can be sharpened by at most $||L_N|| = 0(N^2)$ $(N\to\infty)$, explaining the origin of the difference between the estimate (8.7) and the observed behavior $N^{1/4}$ of the computed L_2-matrix norms.

In the above discussion, we have given numerical evidence for algebraic stability of the Chebyshev-spectral methods for (8.1). We shall now prove rigorously that Chebyshev-spectral methods for (8.1) are algebraically stable.

Proof of Algebraic Stability for Chebyshev-Galerkin Approximation

In the Chebyshev-Galerkin approximation to (8.1), we represent the spectral approximation u_N by the series

$$u_N = \sum_{n=1}^{N} a_n(t)[T_n-(-1)^n T_0] \qquad (8.8)$$

† The condition number of H_N can grow no faster than $N^{5/2}$ as $N\to\infty$. To see this, we note that (5.14) gives $||H_N^{-1}||= 0(N^2)$ while (5.13) and the results that $||\exp(L_N t)|| = 0(N^{1/4})$ for $t\leq 2$ and $||\exp(L_N t)|| \to 0$ as $N\to\infty$ for $t>2$ give $||H_N||= 0(N^{1/2})$ as $N\to\infty$.

Table 8.4

N	Collocation	Tau	Galerkin
10	4.1463×10^2	3.1090×10^2	4.6388×10^2
20	3.0332×10^3	1.2421×10^3	3.2672×10^3
30	9.8746×10^3	2.7938×10^3	1.0464×10^4
40	2.2940×10^4	4.9662×10^3	2.9083×10^4
50	4.4220×10^4	7.7593×10^3	4.6138×10^4

Table 8.4. The condition number $||H_N||\,||H_N^{-1}||$ in the L_2 matrix norm of the Liapounov matrices H_N for the collocation , tau, and Galerkin spectral methods for (8.1-3). For the collocation and Galerkin methods, the condition number seems to grow at most like N^3 as $N\to\infty$, while for the tau method it seems to grow like N^2 as $N\to\infty$.

Recalling (2.34), u_N satisfies

$$\frac{\partial u_N}{\partial t} + \frac{\partial u_N}{\partial x} = \tau_N(t) \sum_{n=0}^{N} \frac{T_n(x)}{c_n} (-1)^n \quad . \tag{8.9}$$

We can determine $\tau_N(t)$ by equating the coefficients of x^N in (8.8):

$$\tau_N(t) = \frac{da_N(t)}{dt} (-1)^N$$

Let us now multiply both sides of (8.9) by $2(1-x)u_N$ and integrate with respect to the Chebyshev weight function $(1-x^2)^{-1/2}$. Thus, the left-hand side of (8.9) becomes

$$2\int_{-1}^{1} (1-x)u_N[\frac{\partial u_N}{\partial t} + \frac{\partial u_N}{\partial x}](1-x^2)^{-\frac{1}{2}}\, dx$$

$$= \frac{d}{dt}\int_{-1}^{1} (1-x)(1-x^2)^{-\frac{1}{2}}\, u_N^2 dx + \int_{-1}^{1} (1-x)^{\frac{1}{2}}(1+x)^{-\frac{1}{2}}\, \frac{\partial u_N^2}{\partial x}\, dx$$

$$= \frac{d}{dt}\int_{-1}^{1} (1-x)(1-x^2)^{-\frac{1}{2}}\, u_N^2 dx + (1-x)^{\frac{1}{2}}(1+x)^{-\frac{1}{2}}\, u_N^2 \Big|_{-1}^{1}$$

$$+ \int_{-1}^{1} u_N^2[\tfrac{1}{2}(1-x)^{-\frac{1}{2}}(1+x)^{-\frac{1}{2}} + \tfrac{1}{2}(1-x)^{\frac{1}{2}}(1+x)^{-\frac{3}{2}}]\, dx \tag{8.10}$$

The boundary term in the last expression vanishes because u_N is a polynomial satisfying $u_N(-1) = 0$. Also,

$$(1-x)u_N = (1-x)\sum_{n=1}^{N} a_n[T_n-(-1)^n T_0] = \sum_{n=1}^{N} a_n[T_n-(-1)^n T_0]$$

$$-\sum_{n=1}^{N} a_n \left[\frac{1}{2} (T_{n+1} + T_{n-1}) - (-1)^n T_1\right]$$

$$= \sum_{n=1}^{N} a_n[T_n-(-1)^n T_0] - \tfrac{1}{2}\sum_{n=1}^{N} a_n(T_{n+1}-(-1)^n T_1] - \tfrac{1}{2}\sum_{n=1}^{N} a_n[T_{n-1}-(-1)^n T_1]$$

(8.11)

The first and third sums on the right in (8.11) are orthogonal to the right side of (8.9). The inner product of $(1-x)u_N$ with the second sum on the right in (8.9) gives

$$-\frac{\pi}{4}(-1)^N \tau_N\, a_N = -\frac{\pi}{4} a_N \frac{da_N}{dt} . \tag{8.12}$$

Combining (8.10) and (8.12), we obtain

$$\tfrac{1}{2}\frac{d}{dt}\int_{-1}^{1}(1-x)(1-x^2)^{-\frac{1}{2}} u_N^2\, dx + \frac{\pi}{8}\frac{d}{dt} a_N^2 \leq 0 \tag{8.13}$$

This inequality proves that u_N is stable in the new norm defined in (8.13):

$$||u||^2 = \int_{-1}^{1} (1-x)(1-x^2)^{-1/2} \; |u(x)|^2 \, dx \; . \tag{8.14}$$

Observe that (8.13) implies that the eigenvalues of L_N have non-positive real parts.

It remains to prove that the norm defined by (8.14) is algebraically equivalent to the usual Chebyshev-L_2 norm. That is, we must show the existence of two functions $c_1(N)$ and $c_2(N)$ such that for every Nth degree polynomial u_N

$$c_1 \int_{-1}^{1} \frac{u_N^2}{\sqrt{1-x^2}} \, dx \le \int_{-1}^{1} \frac{(1-x)u_N^2}{\sqrt{1-x^2}} \, dx \le c_2 \int_{-1}^{1} \frac{u_N^2}{\sqrt{1-x^2}} \, dx \tag{8.15}$$

where $1/c_1(N)$ and $c_2(N)$ grow at most algebraically as $N\to\infty$. The second inequality in (8.15) holds with $c_2(N) = 2$ because $1-x\le 2$.

The first inequality in (8.15) is more difficult to establish. By the mean-value theorem,

$$\int_{-1}^{1} \frac{1-x}{\sqrt{1-x^2}} u_N^2 \, dx = (1-\xi_N) \int_{-1}^{1} \frac{u_N^2}{\sqrt{1-x^2}} \, dx \qquad (-1<\xi_N<1)$$

However this does not prove the required inequality because it is not clear that $1/(1-\xi_N)$ is bounded algebraically as $N\to\infty$ for all polynomials.

To establish the first inequality in (8.15) we use a different approach. We substitute the Chebyshev polynomial expansion

$$u_N = \sum_{n=0}^{N} a_n T_n$$

and obtain

$$\frac{2}{\pi}\int_{-1}^{1}\frac{(1-x)u_N^{\,2}}{\sqrt{1-x^2}}\,dx = 2a_0^2 - 2a_0a_1 + \sum_{n=1}^{N} a_n^{\,2}$$

$$-\frac{1}{2}\sum_{n=2}^{N}(a_n\,a_{n-1} + a_n\,a_{n+1})$$

$$= (a_0\ldots a_N)\quad H_N\quad (a_0\ldots a_N)^T,$$

where H_N is the symmetric, positive definite, $(N+1)\times(N+1)$ tridiagonal matrix whose elements are

$$(H_N)_{jk} = \begin{cases} c_j & \text{if } j = k \\ -\frac{1}{2}c_j & \text{if } j = k-1 \\ -\frac{1}{2}c_k & \text{if } j = k+1 \\ 0 & \text{otherwise,} \end{cases} \tag{8.16}$$

where $c_0 = 2$, $c_n = 1$ if $n > 0$. To complete the demonstration of the first inequality in (8.15), we must show that $H_N \geq c_1(N)I$ where $c_1(N)>0$ and $1/c_1(N)$ is bounded algebraically as $N\to\infty$.

Since H_N is nearly a constant-diagonal tridiagonal matrix, the eigenvalues of H_N can be studied by standard techniques: if $D_N = \det(H_N-\lambda I)$. then D_N satisfies the three-term recurrence relation

$$D_N = (1-\lambda)D_{N-1} - \frac{1}{4}D_{N-2} \quad (N\geq 2). \tag{8.17}$$

Since (8.17) has constant coefficients, it is easy to solve exactly. From this solution, it is not hard to show that the smallest eigenvalue of H_N satisfies

$$\lambda_{\min}^{(N)} \sim \frac{\pi^2}{8N^2} \quad (N\to\infty).$$

Choosing $c_1(N) = \lambda_{\min}(N)$ gives $1/c_1(N) \sim 8N^2/\pi^2$ $(N\to\infty)$. This proves that the norm defined by (8.14) is algebraically equivalent to the Chebyshev norm and, therefore, Chebyshev-Galerkin approximation to (8.1) is algebraically stable. Note also that (8.13) shows that the matrix H_N defined in (8.16) satisfies (5.7b) with $c(N) = 0$. Since $||H_N|| = 0(1)$ and $||H_N^{-1}|| = 0(N^2)$, (5.11) implies that $||\exp(L_N t)|| = 0(N)$ as $N\to\infty$, which also follows directly from (8.15).

We have not yet been able to obtain a rigorous demonstration that $||\exp(L_N t)|| = 0\,(N^{1/4})$ as $N\to\infty$ as found numerically in Table 8.2. Our best result to date is $||\exp(L_N t)|| = 0(N)$ as $N\to\infty$.

Although the problem (8.1) is not well posed in the Chebyshev norm (as shown in Sec. 5), it is well posed in the norm defined by (8.14). Using (8.1) and (8.3), we obtain

$$\int_{-1}^{1} \left(\frac{1-x}{1+x}\right)^{1/2} u\, u_t \, dx = -\int_{-1}^{1}\left(\frac{1-x}{1+x}\right)^{1/2} u\, u_x \, dx$$

$$= -\frac{1}{2} \int_{-1}^{1} u^2 \, (1-x)^{-1/2} \, (1+x)^{-3/2} \, dx \leq 0 .$$

Thus,

$$\frac{d}{dt} \int_{-1}^{1}\left(\frac{1-x}{1+x}\right)^{1/2} u^2 \, dx \leq 0,$$

so that $||e^{Lt}||\leq 1$ in the norm (8.14).

Proof of Algebraic Stability for Chebyshev-Tau Approximation

The proof of algebraic stability for the tau method is similar to that just given for Galerkin approximation. The Chebyshev-tau approximation u_N satisfies

$$\frac{\partial u_N}{\partial t} + \frac{\partial u_N}{\partial x} = \tau_N(t) \; T_N(x) \tag{8.18}$$

$$u_N(-1,t) = 0 \quad ,$$

where

$$u_N = \sum_{n=0}^{N} a_n \, T_n \, . \tag{8.19}$$

Therefore ,

$$(1-x)\ \frac{\partial^2 u_N}{\partial x \partial t} = -N\ \frac{da_N}{dt}\ T_N + \sum_{n=0}^{N-1} b_n\ T_n \qquad (8.20)$$

Moreover, comparing the coefficients of x^N on both sides of (8.18) we find

$$\tau_N(t) = \frac{da_N}{dt} \qquad (8.21)$$

Eqs. (8.18-21) imply

$$\left(\frac{\partial u_N}{\partial t}, (1-x)\frac{\partial^2 u_N}{\partial x \partial t}\right) + \left(\frac{\partial u_N}{\partial x}, (1-x)\frac{\partial^2 u_N}{\partial x \partial t}\right) = -\frac{\pi}{2} N \left(\frac{da_N}{dt}\right)^2 \qquad (8.22)$$

Since

$$\frac{\partial u_N}{\partial t}\Big|_{x=-1} = 0 \ ,$$

we obtain

$$2\left(\frac{\partial u_N}{\partial t}\ ,\ (1-x)\ \frac{\partial^2 u_N}{\partial x \partial t}\right) = \int_{-1}^{1} (1-x)^{1/2}(1+x)^{-1/2}\partial(\partial u_N/\partial t)^2/\partial x\, dx$$

$$= \int_{-1}^{1} (1-x)^{-1/2}(1+x)^{-3/2}(\partial u_N/\partial t)^2\ dx.$$

Therefore ,(8.21) gives

$$\frac{d}{dt}\int_{-1}^{1}(1-x)(1-x^2)^{-\frac{1}{2}}\left(\frac{\partial u_N}{\partial x}\right)^2\ dx \leq 0 \qquad (8.23)$$

This proves that the evolution of $\frac{\partial u_N}{\partial x}$ is stable in the norm (8.14). Finally, the boundedness of $\frac{\partial u_N}{\partial x}$ implies the boundedness of u_N, as will now be shown. If u_N is given by (8.19), then

$$\frac{\partial u_N}{\partial x} = \sum_{n=0}^{N-1} b_n T_n$$

where

$$a_n = \frac{c_{n-1} b_{n-1} - b_{n+1}}{2n} \qquad (n = 1,\dots,N)$$

The boundary condition $u_N(-1,t) = 0$ requires that

$$a_0 = \sum_{n=1}^{N} a_n$$

Therefore, since $\frac{\partial u_N}{\partial x}$ is bounded algebraically as $N\to\infty$, so is u_N.

In Sec. 11 we present a variety of numerical results for the numerical solution of (8.1) by Chebyshev and Legendre spectral methods.

Effect of Boundary Conditions on the Stability of Spectral Methods

Let us discuss the effect of boundary conditions on the stability of the Cheybshev approximations to (8.1). In Sec. 6 it was shown that incorrect treatment of the boundary does not affect the stability (though it does affect the convergence) of the Fourier-Galerkin method. This is not the case for the Chebyshev-spectral methods. Let us assume that we solve (8.1) ignoring the boundary condition (8.3) and suppose that $u_N(x,0) = T_N(x)$. The resulting system of Galerkin equations for $\{a_n\}$ is

$$\frac{\partial a_n}{\partial t} = -\frac{2}{c_n} \sum_{\substack{p=n+1 \\ p+n \text{ odd}}}^{N} p a_p \qquad (8.24)$$

where $a_n(0) = \delta_{nN}$. Eq. (8.24) can easily be solved: $a_{N-k}(t)$ is a polynomial in t of degree k of the form

$$a_{N-k}(t) = \frac{1}{c_{N-k}} (-2)^k \binom{N}{k} t^k + \dots . \qquad (8.25)$$

This solution is clearly not bounded by any finite power of N. Thus, the Chebyshev methods are algebraically unstable when no boundary conditions are applied.

If we had imposed the boundary condition $u(+1,t) = 0$ in addition to, or instead of, the boundary condition $u(-1,t) = 0$, then Chebyshev-spectral solution to (8.1) would be unstable. With $u(+1,t)=0$ instead of (8.3), the Chebyshev-spectral approximations to the operator $-\partial/\partial x$ all have eigenvalues with positive real parts (that grow as $N \to \infty$). Similarly, if we tried to impose the extra boundary condition $\partial u(+1,t)/\partial x = 0$ in addition to $u(-1,t)=0$ [as is frequently done with finite difference methods], an unstable scheme would result.

The effect of imposing $u(+1,t) = 0$ in addition to $u(-1,t) = 0$ is slightly different for Legendre-spectral methods. With $u(-1,t)= u(+1,t)=0$, Legendre-spectral methods for solution of (8.1) are semi-bounded. In fact,

$$(v, [L+L^*] v) = -2 \int_{-1}^{1} v \partial v/\partial x \, dx = 0$$

when $v(\pm 1,t) = 0$, so these methods are semi-bounded and stable.

However, these spectral approximations are not consistent. For example, Galerkin approximation involves expansion of $u(x,t)$ in terms of the functions $\phi_{2n}(x)=P_{2n}(x) - P_0(x)$ $\phi_{2n+1}(x)=P_{2n+1}(x)- P_1(x)$ that satisfy $\phi_n(\pm 1) = 0$. But $\partial u/\partial x$ cannot, in general, be expanded in terms of the functions $\phi_n'(x)$.

The above situations are typical of rapidly converging spectral methods. Spectral methods are extremely sensitive to the proper formulation of boundary conditions. When proper boundary conditions are imposed so the problem is well posed, the methods yield very accurate results; when improper boundary conditions are mistakenly applied, the methods are likely to be explosively unstable. The formulation of stable and convergent spectral methods is strikingly similar to the formulation of well-posed initial-value problems for the continuum equations.

9. Time Differencing

In previous sections we have investigated the properties of spectral approximations to the spatial operator L of the differential equation

$$\frac{\partial u}{\partial t} = Lu .$$

In this section we investigate the properties of time-integration techniques for the solution of the semi-discrete spectral approximations

$$\frac{\partial u_N}{\partial t} = L_N u_N \tag{9.1}$$

Time discretization errors in both finite difference and spectral methods are typically much smaller than are spatial discretization errors. There are two reasons for this: (i) time steps are frequently restricted in size by explicit stability conditions -- stability of the time integration requires that time-differencing errors be small; and (ii) many problems involve several space coordinates so any possible efficiency in the representation of the spatial variation of the dependent variables is quite important to the overall efficiency of the method-- if the number of degrees of freedom necessary to describe a certain three-dimensional field accurately can be reduced by two in each space direction then the total number of degrees of freedom is decreased by a factor 8, but a similar improvement in time differencing gives just a factor 2. We will investigate here only finite-difference methods of finite-order accuracy for timewise solution of (9.1) despite the infinite-order accuracy in space of many of the spectral methods discussed in earlier sections. No efficient, infinite-order accurate time-differencing methods for variable coefficient problems are yet known. The current state-of-the-art of time-integration techniques for spectral methods is far from satisfactory on both theoretical and practical grounds and the results to be presented here must be regarded as only a beginning.

One of our prime goals is to investigate the <u>stability</u> of time differencing methods for the solution of (9.1). To do this we must first explain how to extend the stability definitions given in Sects. 4 and 5. Let $u_N^n(x) = \hat{u}_N(x,n\Delta t)$ be the approximation to the solution of (9.1) at time $n\Delta t$, where Δt is a time step. Time differencing methods involve approximating (9.1) in some way to give a rule for constructing u_N^{n+1}:

$$u_N^{n+1} = K_N(\Delta t)u_N^n , \tag{9.2}$$

where K_N is an operator acting on u_N. Using this rule repetitively it follows that

$$\hat{u}_N(x,n\Delta t) = [K_N(\Delta t)]^n u_N(x,0), \tag{9.3}$$

where, for notational simplicity, we assume Δt fixed. We say that (10.2) is <u>strongly</u> stable if

$$||[K_N(\Delta t)]^n|| \le K(n\Delta t) \tag{9.4}$$

for all N and n sufficiently large and Δt sufficiently small. Here $K(T)$ is a finite function of T. We define generalized stability by replacing $K(T)$ in (9.4) by $N^{r+sT}K(T)$ as in (5.2).

A sufficient, though not necessary, condition for strong stability (9.4) is

$$||K_N(\Delta t)|| - 1 \leq \kappa\Delta t \tag{9.5}$$

for some finite κ and all Δt sufficiently small. If $K_N(\Delta t)$ is a normal matrix then strong stability is assured in the L_2 matrix norm if the eigenvalues λ of K_N satisfy the von Neumann condition

$$\max|\lambda| \leq 1 + \kappa\Delta t \tag{9.6}$$

for sufficiently small Δt. If K_N is not a normal matrix, then (9.6) is still a necessary, though not sufficient, condition for stability in the sense of (9.4).

The importance of these stability definitions is that they lead to the fully discrete form of the equivalence theorem (see Sec. 4): a scheme is consistent if

$$||(\frac{K_N(\Delta t) - I}{\Delta t} - L)u|| \to 0 \tag{9.7}$$

as $N \to \infty$ and $\Delta t \to 0$ for all u in a dense subspace of H; a scheme is convergent if

$$||u_N^n - u(n\Delta t)|| \to 0$$

as $N \to \infty$ and $\Delta t \to 0$ for all n satisfying $0 \leq n\Delta t \leq T$ and all $u(0) \in \mathcal{H}$. The equivalence theorem states that for consistant approximations to well-posed problems, stability is equivalent to convergence.

Let us now study the stability properties of some specific time-differencing methods.

<u>Implicit time-integration methods</u>

Two time-integration methods that are unconditionally stable in the generalized sense for every algebraically stable spectral method are the Crank-Nicolson scheme and the backwards Euler scheme. For any semi-discrete spectral approximation (9.1) to $u_t = Lu$, the Crank-Nicolson time-differencing scheme is given by

$$u_N^{n+1} - u_N^{n} = \Delta t \; L_N\left(\frac{u_N^{n+1} + u_N^{n}}{2} \right) \tag{9.8}$$

and the backwards Euler scheme is given by

$$u_N^{n+1} - u_N^{n} = \Delta t \, L_N \, u_N^{n+1} \, . \tag{9.9}$$

Let us prove that (9.8) and (9.9) are stable in the generalized sense. If (9.1) is algebraically stable there exists a family of positive definite Hermitian matrices $\{H_N\}$ such that

$$H_N L_N + L_N^* H_N \leq \alpha(N) \, H_N$$

or, equivalently,

$$H_N^{1/2} L_N H_N^{-1/2} + H_N^{-1/2} L_N^* H_N^{1/2} \leq \alpha(N) I,$$

where $\alpha(N) \leq d \ln N$ for some finite d. Substituting

$$v_N^{\ n} = H_N^{\ 1/2} u_N^{\ n}$$

into (9.8-9), we obtain, respectively,

$$v_N^{\ n+1} - v_N^{\ n} = \Delta t\ M_N \left(\frac{v_N^{\ n+1} + v_N^{\ n-1}}{2} \right), \tag{9.10}$$

$$v_N^{\ n+1} - v_N^{\ n} = \Delta t\ M_N\, v_N^{\ n+1}, \tag{9.11}$$

where

$$M_N = H_N^{\ 1/2} L_N H_N^{\ -1/2}.$$

Taking the scalar product of (9.10) with $v_N^n + v_N^{n+1}$, we get

$$\|v_N^{\ n+1}\|^2 - \|v_N^{\ n}\|^2 = \frac{\Delta t}{2}\left((v_N^{\ n+1} + v_N^{\ n}), \left(\frac{M_N + M_N^{\ *}}{2}\right)(v_N^{n+1} + v_N^{n}) \right)$$

$$\leq \frac{\alpha \Delta t}{4} \|v_N^{\ n+1} + v_N^{\ n}\|^2 \leq \frac{\alpha \Delta t}{2} [\|v_N^{\ n+1}\|^2 + \|v_N^{\ n}\|^2] \tag{9.12}$$

Therefore,

$$||v_N^{n+1}||^2 \leq \frac{(1+\frac{1}{2}\alpha\Delta t)}{(1-\frac{1}{2}\alpha\Delta t)}||v_N^{n}||^2, \tag{9.13}$$

which proves generalized stability for v_N and, hence, also for

$$u_N = H_N^{-1/2} v_N.$$

Similarly, we may show that the backwards Euler method is unconditionally stable in the generalized sense. Taking the scalar product of (9.11) with $v_N^{n+1} + v_N^n$ gives

$$\begin{aligned} ||v_N^{n+1}||^2 - ||v_N^{n}||^2 &= \Delta t\ (M_N v_N^{n+1}, v_N^{n} + v_N^{n+1}) \\ &= \Delta t\ (M_N v_N^{n+1}, 2v_N^{n+1} - \Delta t M_N v_N^{n+1}) \\ &\leq \alpha\Delta t\ ||v_N^{n+1}||^2, \end{aligned} \tag{9.14}$$

so that

$$||v_N^{n+1}||^2 \leq \frac{1}{1-\alpha\Delta t}||v_N^n||^2 ,$$

proving generalized stability of v_N and, hence, u_N.

Note that the above proofs show that if $\alpha(N)$ is a bounded function of N then $v_N = H_N^{1/2} u_N$ is strongly stable for both the Crank- Nicolson and backwards Euler schemes.

Spectral approximations using Fourier series

Next, we consider several time integration methods for Fourier series spectral approximations to

$$u_t + u_x = 0$$

with periodic boundary conditions. As shown in Sec. 6, the collocation equations are

$$\frac{\partial u_N}{\partial t} = C^{-1}DCu_N \tag{9.15}$$

where the 2N x 2N matrices C and D are defined in (6.3).

The 'leapfrog' time differencing approximation to (9.15) is the explicit two-level scheme

$$u_N^{n+1} - u_N^{n-1} = 2\Delta t\ C^{-1}\ DCu_N^n \tag{9.16}$$

Thus, in the leapfrog scheme

$$K_N(\Delta t)u_N^n = u_N^{n-1} + 2\Delta tC^{-1}DCu_N^n ,$$

so K_N is a two-level evolution operator since it depends on both u_N^{n-1} and u_N^n. The definitions of stability, convergence, and consistency given above extend easily to this case.

We shall show that (9.16) is strongly stable provided that

$$\Delta t < \frac{1}{2\pi(N-1)} \tag{9.17}$$

To show this we first recall from Sec. 6 that C is unitary and D is skew-Hermitian. Therefore, $A = C^{-1}DC$ is also skew-Hermitian, and hence normal, so that

$$||A|| = 2\pi(N-1) \ .$$

Now we take the inner product of (9.16) with $u_N^{n+1} + u_N^{n-1}$ to get

$$||u_N^{n+1}||^2 - ||u_N^{n-1}||^2 = 2\Delta t \mathrm{Re}(u_N^{n+1} + u_N^{n-1}, Au_N^n) \ ,$$

since $u_N^{n\pm1}$ and u_N^n are real. Since $A^* = -A$, we obtain

$$U_N^n \equiv ||u_N^{n+1}|| + ||u_N^n||^2 - 2\Delta t \mathrm{Re}(u_N^{n+1}, Au_N^n)$$

$$= ||u_N^n||^2 + ||u_N^{n-1}|| - 2\Delta t \mathrm{Re}(u_N^n, Au_N^{n-1}) \equiv U_N^{n-1}$$

so $U_N^n = U_N^0$. Schwarz' inequality implies that

$$|\mathrm{Re}(u_N^{n+1}, Au_N^n)| \leq ||A|| \ ||u_N^{n+1}|| \ ||u_N^n||$$

so that if (9.17) is satisfied, i.e. $\Delta t||A|| \le 1-\varepsilon$ for some $\varepsilon > 0$,

$$|2\Delta t \mathrm{Re}(u_N^{n+1}, Au_N^n)| \le 2(1-\varepsilon)\ ||u_N^{n+1}||\ ||u_N^n|| \ .$$

Using this result, we obtain

$$\varepsilon(||u_N^{n+1}|| + ||u_N^n||^2) + (1-\varepsilon)\ (||u_N^{n+1}|| - ||u_N^n||)^2 \le U_N^n = U_N^0 \tag{9.18}$$

Since U_N^0 is a bounded function of N (because of the smoothness of the initial conditions), we see that $||u_N^{n+1}||$ is bounded for all N and n, proving strong stability.

Another way to prove that the leapfrog and Crank-Nicolson time differencing schemes are strongly stable for (9.15) is to use a modal analysis, which is justified because A is normal. Thus, if u_N^0 is an eigenfunction of A with eigenvalue λ, the Crank-Nicolson approximation to $K_N(\Delta t)$ is

$$K_N(\Delta t) u_N^0 = (1 + \tfrac{1}{2}\lambda\Delta t)/(1 - \tfrac{1}{2}\lambda\Delta t)\ u_N^0 \tag{9.19}$$

Since the eigenvalues λ of $C^{-1}DC$ are all pure imaginary, it follows that $||K_N(\Delta t)|| = 1$, so Crank-Nicolson differencing is stable.

Still another time differencing method for solution of (9.15) is to use a Runge-Kutta scheme. It easily verified the first and second-order Runge-Kutta methods are unstable unless Δt satisfies conditions that are much more restrictive than (9.17). With the first-order Euler method

$$u_N^{n+1} = u_N^n + \Delta t A u_N^n ,$$

stability requires that $N^2\Delta t$ be bounded as $\Delta t \to 0$ [because $||K_N(\Delta t)|| = 1 + O(N^2 \Delta t^2)$] ; with the second-order scheme

$$\tilde{u}_N^{n+1/2} = u_N^n + \frac{1}{2}\Delta t A u_N^n$$

$$u_N^{n+1} = u_N^n + \Delta t A \tilde{u}_N^{n+1/2} ,$$

stability requires that $N^{4/3}\Delta t$ be bounded as $\Delta t \to 0$. However, the third and fourth-order Runge-Kutta methods give conditional stability restrictions like (9.17) which we will now derive.

The third-order Runge-Kutta scheme may be written for a linear equation like (9.1) as

$$u_N^{n+1} = [I + \Delta t A + 1/2(\Delta t A)^2 + 1/6(\Delta t A)^3]u_N^n = K_N(\Delta t)u_N^n. \tag{9.20}$$

Since $K_N(\Delta t)$ given by (9.20) is normal,

$$||K_N(\Delta t)|| = \max_\lambda |1 + \lambda\Delta t + 1/2(\lambda\Delta t)^2 + 1/6(\lambda\Delta t)^3|,$$

where the maximum is taken over all the eigenvalues of A. eigenvalues of A are ik with $|k| \le 2\pi(N-1)$, so (9.6) is satisfied provided that

$$\Delta t < \frac{\sqrt{3}}{2\pi(N-1)} . \tag{9.21}$$

Thus, this method allows time steps that can be $\sqrt{3}$ times larger than with the leapfrog scheme while maintaining stability. However, if the operator A is complicated, the third-order Runge-Kutta scheme requires about 3 times as much work as leap-from at each time step, so it is probably not competitive.

Similar analysis of the fourth-order Runge-Kutta scheme gives the stability condition

$$\Delta t < \frac{\sqrt{2}}{\pi(N-1)} \tag{9.22}$$

Thus time steps can be nearly three times larger than with leapfrog steps. However, fourth-order Runge-Kutta differencing requires about four times the work of leapfrog differencing, so the scheme is probably not too useful unless very high accuracy is desired.

Now we shall consider time-differencing methods for Fourier series spectral approximations to the heat equation with periodic boundary conditions:

$$u_t = u_{xx} \qquad (0 \leq x \leq 1) \tag{9.23}$$

Collocation using Fourier series gives the spectral equations

$$\frac{\partial u_N}{\partial t} = C^{-1}D^2Cu_N \tag{9.24}$$

The matrix $C^{-1} D^2 C$ is negative definite. Because (9.19) still holds and all eigenvalues λ are negative, Crank-Nicolson time differencing is unconditionally stable. On the other hand, it is easy to show that leapfrog differencing is unconditionally unstable. In fact, if u_N^0 is an eigenfunction of $C^{-1}D^2C$ with eigenvalue $\lambda < 0$ then $||K_N(\Delta t)^n u_N^0||$ grows like $(-\lambda\Delta t + \sqrt{1+(\lambda\Delta t)^2})^n \sim e^{-\lambda(n\Delta t)}$ as $\Delta t \to 0$ for fixed λ and $n\Delta t$. Since $\max|\lambda| = 4\pi^2(N-1)^2$ grows like N^2 as $N \to \infty$, $||K_N(\Delta t)^n u_N^0||$ cannot be bounded be a finite function of $n\Delta t$ for all N, proving unconditional instability.

Another way to solve (9.24) is to use a generalized Dufort-Frankel scheme

$$\frac{u_N^{n+1} - u_N^{n-1}}{2\Delta t} = C^{-1}D^2Cu_N^n - \gamma N^2(u_N^{n+1} - 2u_N^n + u_N^{n-1}) \tag{9.25}$$

If $\gamma \geq \pi^2$ then this method is unconditionally stable (Gottlieb & Gustaffson 1976).

Similarly, Euler time differencing of (9.24) is conditionally stable. Stability requires that

$$\Delta t \max|\lambda| \leq 2 \text{ or } \Delta t \leq [2\pi^2(N-1)^2]^{-1}. \tag{9.26}$$

Higher-order Adams-Bashforth schemes have similar conditional stability limits.

Time-differencing for mixed initial-boundary value problems

Some care is necessary in the formulation of time-differencing methods for spectral approximations to mixed initial-boundary value problems. The sensitivity of spectral methods to the proper formulation of boundary conditions, as shown in Sects. 6-8, carries over to the formulation of time-differencing methods for these approximations. For example, for most mixed initial-boundary value problems leap-frog time differencing is unconditionally unstable for spectral approximations. Furthermore, explicit time integration methods may be unduly restricted by conditional stability requirements in spectral approximations. The origin of these severe restrictions is the very high resolution of spectral methods near boundaries. Thus, it is frequently necessary to combine special kinds of implicit time-integration methods with spectral approximations in order to maintain high accuracy at reasonable computational cost. Several examples will be given later.

Let us begin by studying time-differencing methods for the Chebyshev-spectral approximation to the mixed initial-boundary value problem (8.1-3);

$$u_t + u_x = 0 \quad (-1\le x\le 1,\ t>0), \tag{9.27}$$

$$u(x,0) = f(x) \quad (-1\le x\le 1), \tag{9.28}$$

$$u(-1,t) = 0 \quad (t>0). \tag{9.29}$$

In Sec. 8, we proved that various semi-discrete spectral approximations to (9.27-29) are algebraically stable.

Let us first consider the leapfrog time-differencing scheme

$$u_N^{n+1} = u_N^{n-1} + 2\Delta t\, L_N\, u_N^{n}, \tag{9.30}$$

where $u_N^n(x)$ is the time-discretized approximation to $u_N(x, n\Delta t)$, Δt is the time step, and the semi-discrete approximation is $\partial u_N/\partial t = L_N u_N$.
This scheme is unconditionally unstable for any Δt as $N\to\infty$. To show this we recall that in Sec. 8. we proved that the eigenvalues of L_N have negative real part (see Table 8.3) and that the largest eigenvalue of L_N has a negative real part that grows like N^2 as $N\to\infty$. Let us rewrite (9.30) in the 2 x 2 block-matrix form

$$\begin{pmatrix} u_N^{n+1} \\ u_N^{n} \end{pmatrix} = \begin{pmatrix} 2\,\Delta t\, L_N & I \\ I & 0 \end{pmatrix} \begin{pmatrix} u_N^{n} \\ u_N^{n-1} \end{pmatrix} \tag{9.31}$$

If the eigenvalues of L_N are denoted as μ_N, then the eigenvalues of the matrix on the right in (9.31) are

$$\lambda_N^{(\pm)} = \mu_N\, \Delta t \pm \sqrt{1 + (\Delta t)^2\, \mu_N^{\;2}} \tag{9.32}$$

For fixed N and $\Delta t \to 0$,

$$\lambda_N^{(-)} = e^{-\mu_N \Delta t} \, (1 + 0(\Delta t^2)). \tag{9.33}$$

Thus

$$\left[\lambda_N^{(-)}\right]^n = (-1)^n \, e^{-\mu_N n \Delta t} \, (1 + 0(\Delta t)) \quad (0 \leq n\Delta t \leq T, \Delta t \to 0) \tag{9.34}$$

Since $||K_N(\Delta t)^n|| \geq |\lambda_N^{(-)}|^n$ and there are eigenvalues of L_N with negative real part of order N^2, no inequality of the form (9.4) can be satisfied. Thus, leapfrog time differencing of the Chebyshev approximations to (9.27-29) is unconditionally unstable.

There are several conditionally stable explicit time-differencing approximations that can be used with spectral approximations to (9.27-29). Two examples are the Adams-Bashforth scheme

$$u_N^{n+1} = u_N^n + \frac{3}{2} \Delta t \, L_N^n = \frac{1}{2} \Delta t \, L_N^{n-1} \tag{9.35}$$

and the modified Euler scheme

$$\hat{u}_N^{n+1} = u_N^n + \Delta t \, L_N \, u_N^n \tag{9.36a}$$

$$u_N^{n+1} = u_N^n + \frac{1}{2} \Delta t \, L_N \, u_N^n + \frac{1}{2} \Delta t \, L_N \, \hat{u}_N^{n+1} \tag{9.36b}$$

The modified Euler scheme (9.36) is in practice stable provided the stability condition

$$\Delta t \leq \frac{8}{N^2} \tag{9.37}$$

is satisfied. A similar stability condition holds for the Adams-Bashforth scheme.

The fact that the stability limit in (9.37) depends on $1/N^2$ rather than $1/N$ is not very surprising because the Chebyshev collocation points $\{\cos \pi n/N;\ n=0,1,\ldots,N\}$ are spaced by a distance of order $1/N^2$ near the boundaries. Since the wave speed in (9.27) is 1 the wave propagates from one grid point to the next in a time of order $1/N^2$ so time steps must be smaller than this to maintain explicit stability.

The explicit stability restriction (9.37) for Chebyshev-spectral methods with N polynomials should be contrasted with the corresponding stability conditions for finite difference approximations to (9.27-29). With N gridpoints uniformly spaced in the interval $-1 \leq x \leq 1$, the grid spacing is $2/N$ so the Courant stability condition is $\Delta t \leq 2/N$. As $N \to \infty$, this stability condition on finite difference schemes is much weaker than the condition (9.37) on the spectral approximations. A semi-implicit technique that permits stable time-differencing with spectral methods with a stability condition like that of finite-difference schemes will be discussed later in this section.

In order to prove that the modified Euler method (9.36) is stable, we begin by noting that (9.36) is equivalent to the second-order Taylor series method

$$u_N^{n+1} = (I + \Delta t\ L_N + \tfrac{1}{2}(\Delta t)^2\ L_N{}^2)\ u_N^n \equiv G_N\ u_N^n \tag{9.38}$$

A sufficient condition for algebraic stability of (9.38) is the existence of positive-definite symmetric matrices S_N such that

$$G_N^T\ S_N\ G_N \le S_N \tag{9.39a}$$

and the condition number of S_N satisfies

$$||S_N||\,||S_N^{-1}|| = O(N^\beta) \quad (N\to\infty) \tag{9.39b}$$

for some finite β. If (9.39) holds then

$$(G_N^T)^n\ S_N\ (G_N)^n \le (G_N^T)^{n-1}\ S_N\ (G_N)^{n-1} \le \ldots \le S_N$$

so that

$$S_N^{-1/2}\ (G_N^T)^n\ S_N^{1/2}\ S_N^{1/2}\ (G_N)^n\ S_N^{-1/2} \le I.$$

Therefore,

$$||S_N^{1/2}\ (G_N)^n\ S_N^{-1/2}|| \le 1,$$

so that

$$||u_N^n|| = ||(G_N)^n u_N^0|| \le ||S_N^{-1/2}|| \, ||S_N^{1/2}(G_N)^n S_N^{-1/2}||$$

$$\times \, ||S_N^{1/2}|| \, ||u_N^0|| = 0 \, (N^\beta ||u_N^0||) \quad (N\to\infty).$$

To complete the stability proof we must investigate under what conditions matrices S_N satisfying (9.39) exist. One choice for S_N is just the Liapounov matrices of L_N; these matrices satisfy

$$S_N L_N + L_N^T S_N = - I \tag{9.40}$$

From Table 8.4 we observe that the Liapounov matrices for spectral approximations to (9.27-29) have algebraically bounded condition number. Using (9.38), we obtain

$$G_N^T S_N G_N = [I + \Delta t \, L_N^T + \tfrac{1}{2} (\Delta t)^2 (L_N^2)^T] \, S_N \, [I+\Delta t L_N + \tfrac{1}{2}(\Delta t)^2 (L_N)^2]$$

or

$$\begin{aligned} G_N^T S_N G_N = {} & S_N + \Delta t \, (L_N^T S_N + S_N L_N) \\ & + \tfrac{1}{2}(\Delta t)^2 \, [(L_N^2)^T S_N + 2L_N^T S_N L_N + S_N L_N^2] \\ & + \tfrac{1}{2} (\Delta t)^3 [(L_N^2)^T S_N L_N + L_N^T S_N L_N^2] + \tfrac{1}{4} (\Delta t)^4 (L_N^2)^T S_N L_N^2 . \end{aligned}$$

From (9.40), it follows that

$$(L_N^2)^T S_N + L_N^T S_N L_N = -L_N^T$$

$$L_N^T S_N L_N + S_N L_N^2 = -L_N$$

$$(L_N^2)^T S_N L_N + L_N^T S_N L_N^2 = -L_N^T L_N$$

so that

$$G_N^T S_N G_N = S_N - \Delta t I - \tfrac{1}{2}(\Delta t)^2 \, [L_N^T + L_N]$$

$$- \frac{1}{2} (\Delta t)^3 L_N^T L_N + \frac{1}{4} (\Delta t)^4 (L_N^2)^T S_N L_N^2$$

Thus, (9.39a) is satisfied provided that

$$-\Delta t \, (L_N^T + L_N) \leq 2I \tag{9.41}$$

$$\Delta t \, L_N^T S_N L_N \leq 2I \tag{9.42}$$

If (9.41-42) are satisfied then the modified Euler method for (9.27-29) is algebraically stable.

At first, it may appear that the stability condition (9.42) is much more severe than the stability condition (9.41). The Liapounov matrices of Chebyshev polynomial approximations to the wave equation satisfy $||S_N|| \geq O(1)$ as $N \to \infty$ while the operator L_N satisfies $||L_N|| = O(N^2)$ [see Sec. 8], so that (9.4 2) seems to require that $\Delta t = O(1/N^4)$ as $N \to \infty$. However, the stability condition (9.42) is no more restrictive than the stability condition (9.41) To see this we use (9.40) written in the form

$$L_N^T S_N L_N L_N^{-1} + (L_N^T)^{-1} L_N^T S_N L_N = - I$$

to obtain the representation [see (5.13)]

$$L_N^T S_N L_N = \int_0^\infty \exp[(L_N^{-1})^T t] \exp[L_N^{-1} t] \, dt \ . \tag{9.43}$$

It may be shown that the norm of the integrand of (9.43) is $O(1)$ as $N \to \infty$ for $t = O(N^2)$ and that the norm decays rapidly to zero as $t \to \infty$. Therefore,

$$||L_N^T S_N L_N|| = 0 \ (N^2) \quad (N \to \infty) \tag{9.44}$$

showing that the stability condition (9.42) is of the form $\Delta t = O(1/N^2)$.

Semi-implicit methods

When explicit time-stepping methods are used to solve semi-discrete spectral equations for the hyperbolic problem

$$\frac{\partial u}{\partial t} + a(x)\,\frac{\partial u}{\partial x} = 0 \quad (-1 \le x \le 1) \tag{9.45}$$

with appropriate boundary conditions [that depend on the sign of a(x)], there result stability conditions of the form

$$\Delta t \lesssim \min\left\{\frac{1}{N^2|a(1)|},\ \frac{1}{N^2|a(-1)|},\ \frac{1}{N \max\limits_{|x|\le 1} |a(x)|}\right\} \tag{9.46}$$

These stability limits can be derived heuristically from the Courant stability condition

$$\Delta t < \frac{\Delta x_{eff}}{|a_{eff}|} \tag{9.47}$$

where a_{eff} is the effective wave propagation speed in a direction in which there is effective grid resolution Δx_{eff}. Near the boundaries $x=\pm 1$, the Chebyshev-spectral methods have resolution $\Delta x_{eff} = 0(1/N^2)$ as $N\to\infty$ while $a_{eff} = a(\pm 1)$; in the interior of $-1<x<1$, Chebyshev series have effective resolution $\Delta x_{eff} = 0(\frac{1}{N})$ as $N\to\infty$ while the largest wave spped is $\max|a(x)|$. Thus, (9.47) implies (9.46) for the Chebyshev-spectral methods.

The stability condition (9.46) is too severe for many applications because it requires that $\Delta t = 0(1/N^2)$.

In order to relax this severe constraint, we use a semi-implicit method in which the propagation through the high-resolution boundary is treated implicitly, but the propagation through the interior is treated explicitly.

One possible semi-implicit scheme is the following two-step method. Let L_N be the Chebyshev-spectral approximation to $-a(x)\ \frac{\partial}{\partial x}$ with appropriate boundary conditions applied, and L_N^+, L_N^- be the Chebyshev spectral approximations to the constant coefficient wave operators $-a(+1)\partial/\partial x$, $-a(-1)\partial/\partial x$, respectively, again with appropriate boundary conditions applied. A semi-implicit two-step scheme is given by

$$u_N^{n+\frac{1}{2}} - \tfrac{1}{2}\Delta t\, L_N^-\, u_N^{n+\frac{1}{2}} = u_N^n + \tfrac{1}{2}\Delta t\, (L_N - L_N^-)\, u_N^n \tag{9.48a}$$

$$u_N^{n+1} - \tfrac{1}{2}\Delta t\, L_N^+\, u_N^{n+1} = u_N^{n+\frac{1}{2}} + \tfrac{1}{2}\Delta t\, (L_N - L_N^+)\, u_N^{n+\frac{1}{2}} \tag{9.48b}$$

The scheme (9.48) is stable if the stability condition

$$\Delta t \lesssim \frac{1}{N \max |a(x)|} \tag{9.49}$$

is satisfied.

The condition (9.49) is sufficient to ensure stability, but the semi-implicit scheme (9.48) may be stable even if (9.49) is violated. If $\max|a(x)| < |a(1)|$ or $\max|a(x)| < |a(-1)|$, (9.48) is usually unconditionally stable for sufficiently large N (see Sec. 8 of Orszag 1974). The implementation of (9.48) on a computer is straightforward and efficient; the properties of Chebyshev polynomials summarized in the Appendix show that the implicit equations (9.48) are essentially tridiagonal matrix equations.

The reason that the semi-implicit method outlined above does not have a stability restriction like $\Delta t = O(1/N^2)$ can be understood as follows. By subtracting L_N^+ and L_N^- in succeeding half time-steps, the explicit part of the calculation is similar to that in solving an equation of the form

$$\frac{\partial u}{\partial t} + (1-x^2)\ b(x)\ \frac{\partial u}{\partial x} = 0 \tag{9.50}$$

where the wave speed vanishes at $x = \pm 1$. If $b(x) = b$, a constant, the Chebyshev-tau equations for (9.50) are just

$$\frac{da_n}{dt} = 2\,\frac{1}{c_n}\, b\ [(n-1)\ a_{|n-1|} - (n+1)\ a_{n+1}\ ] \tag{9.51}$$

where $c_0 = 2$ and $c_n = 1$ for $n>0$. By Gerschgorin's theorem, $||L_N||$ for (9.51) satisfies

$$||L_N|| = O(bN) \qquad (N\to\infty), \tag{9.52}$$

so the explicit time step restriction is $\Delta t = O(1/bN)$ as $N\to\infty$.

We note that Chebyshev-spectral approximations to (9.50) are stable when no boundary conditions are applied. In fact, using Galerkin approximation and the Chebyshev inner product, we obtain

$$(u_N, \frac{\partial u_N}{\partial t} + b(1-x^2) \frac{\partial u_N}{\partial x}) = 0 .$$

so

$$\frac{d}{dt} \int_{-1}^{1} \frac{u_N^{\,2}}{\sqrt{1-x^2}} dx = -b \int_{-1}^{1} \sqrt{1-x^2} \frac{\partial}{\partial x} u_N^{\,2} dx$$

$$= -b \int_{-1}^{1} \frac{x u_N^{\,2}}{\sqrt{1-x^2}} dx \leq |b| \int_{-1}^{1} \frac{u_N^{\,2}}{\sqrt{1-x^2}} dx .$$

Therefore,

$$||u_N(t)||^2 < e^{|b|t} ||u_N(0)||^2 .$$

proving stability.

There are other attractive semi-implicit schemes for (9.45). For example, suppose $a(x)$ is one-signed, say $a(x)>0$, and let $a_{max} = \max a(x)$. Define L_N^{max} as the Chebyshev approximation to $-\frac{1}{2} a_{max} \frac{\partial}{\partial x}$ with boundary conditions imposed at $x = -1$. A semi-implicit Chebyshev spectral scheme for (9.45) is

$$u_N^{n+1} - \Delta t\, L_N^{max} u_N^{n+1} = u_N^n + \Delta t (L_N - L_N^{max}) u_N^n . \tag{9.53}$$

The scheme (9.53) is usually unconditionally stable and avoids the severe time step restriction (9.46). It is also easy to implement efficiently because L_N^{max} is a Chebyshev approximation to a constant-coefficient wave operator.†

The same kind of trick stabilizes spectral methods for non-linear equations. For example, if we are solving the equation

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0$$

during a time interval in which $u(x,t)$ is smooth (no shock waves), then we may use the semi-implicit scheme

$$\frac{\partial u}{\partial t} + \frac{1}{2}u_{max} \frac{\partial u}{\partial x} = \left(\frac{1}{2} u_{max} - u\right) \frac{\partial u}{\partial x}$$

in which the terms on the left are treated implicitly in time, while those on the right are treated explicitly. Here u_{max} is an estimate of the largest value of $u(x,t)$. Similar semi-implicit methods are effective in eliminating (or at least relaxing) time-step restrictions for finite-difference methods. The key idea is to recognize the source term of a numerical instability and then to approximate it by a simple expression that can easily be treated implicitly.

---

† D. Haidvogel has pointed out that the semi-implicit scheme (9.53) with L_N^{max} replaced by a Chebyshev spectral approximation to $\frac{1}{2}(bx+c)\partial/\partial x$, where $b+c = a(+1)$, $c-b = a(-1)$, is also stable under the weak restriction (9.49). The resulting implicit equations are still tridiagonal [see (A.9), (A.18)].

Several other examples of semi-implicit methods should make the general technique clear. For the variable coefficient heat equation

$$u_t = k(x)\ u_{xx} \qquad (-1 \le x \le 1)$$

with suitable boundary conditions at $x = \pm 1$ and $k(x) > 0$, Chebyshev-spectral methods give explicit time-step stability conditions of the form

$$\Delta t \lesssim \min\left\{ \frac{1}{k(-1)N^4}, \frac{1}{k(1)N^4}, \frac{1}{N^2 \max\limits_{|x|<1} k(x)} \right\} \tag{9.54}$$

The very severe time step restriction that $\Delta t = O(1/N^4)$ as $N \to \infty$ is due to the high resolution of Chebyshev series near the boundaries $x = \pm 1$. To avoid this problem we can use a semi-implicit method. Let L_N be the Chebyshev-spectral approximation to $k(x)\partial^2/\partial x^2$ and let L_N^{max} be the Chebyshev-spectral approximation to $\frac{1}{2}k_{max}\ \partial^2/\partial x^2$ where $k_{max} = \max k(x)$. The semi-implicit scheme (9.53) with L_N^{max} defined in this way seems to be unconditionally stable (Orszag 1974) and certainly does not have any stability restrictions of the form (9.54).

Finally, we comment on the need for implicit or semi-implicit schemes in multi-dimensional problems. If we wish to solve the Navier-Stokes equations

$$\frac{\partial \vec{u}}{\partial t} + \vec{u} \cdot \vec{\nabla} \vec{u} = - \nabla p + \nu \nabla^2 \vec{u}$$

$$\nabla \cdot \vec{u} = 0 \tag{9.55}$$

for incompressible fluid flow, the treatment of the various terms should be guided closely by the type of stability restrictions they impose.

If $\nu = 0$ then we need only consider the types of stability restrictions induced by the advective term $-\vec{u}.\nabla\vec{u}$ and by the pressure term $-\vec{\nabla}p$; we will not discuss the effect of the pressure because it is closely connected to the incompressiblity condition $\nabla.\vec{u}=0$ and is not relevant to the semi-implicit ideas discussed here. At a boundary of the flow, it is appropriate to specify boundary conditions on $\vec{u}.\vec{n}$ where $\vec{n}$ is the normal to the boundary. If the boundary is solid and stationary, then $\vec{u}.\vec{n}=0$ and we are in a situation similar to that modelled by (9.50). The effective convective speed normal to the boundary vanishes, so spectral methods exhibit no unusual time stepping restrictions. However, if fluid is being blown into or sucked out of the boundary so $\vec{u}.\vec{n} \neq 0$, then semi-implicit methods must be applied to avoid unreasonably restrictive conditions like (9.46) on the time steps.

If $\nu > 0$, then the viscous terms in the Navier-Stokes equations should be treated implicitly to avoid unreasonable time step restrictions due to the high resolution of spectral approximations near the boundary.

10. Efficient Implementation of Spectral Methods

There are two aspects of the efficient implementation of spectral methods that we discuss here: (i) evaluation of derivatives; (ii) evaluation of nonlinear and nonconstant coefficient terms; (iii) roundoff errors. More details on these matters are given elsewhere (see the References).

Evaluation of derivatives

An efficient procedure to obtain the expansion coefficients of derivatives of a function $f(x)$ in terms of the expansion coefficients of $f(x)$ is to use recurrence relations. For example, to evaluate the term

$$S_n = \sum_{\substack{p=n+1 \\ p+n \text{ odd}}}^{N} p\, a_p$$

that appears in the Chebyshev equations (2.11), (2.19), and (2.32), we use the recurrence

$$S_n = S_{n+2} + (n+1)a_{n+1} \qquad (0 \le n \le N-1) \tag{10.1}$$

with $S_N = S_{N+1} = 0$. In this way, S_n is evaluated for all n using only N arithmetic operations. The existence of the recurrence relation (10.1) is ensured by the recurrence property

$$2T_n = \frac{T'_{n+1}}{n+1} - \frac{T'_{n-1}}{n-1} \qquad (n>1)$$

satisfied by the Chebyshev polynomials. Similarly, it is possible to derive recurrence relations to evaluate efficiently the coefficients of arbitrary derivatives of functions expanded in Chebyshev and other classical polynomial expansions.

<u>Evaluation of nonlinear and nonconstant coefficient terms</u>

The most efficient way to evaluate nonlinear and general nonconstant terms in spectral approximations is to apply transform methods. The key idea is to apply fast Fourier transforms and other transforms to transform efficiently between spectral representations of a function $f(x)$ and physical-space representations of $f(x)$. With Chebyshev polynomial expansions, fast Fourier transforms permit the evaluation of arbitrary nonlinear and nonconstant coefficients terms in order $N \log N$ arithmetic operations.

In general, collocation methods give approximations to nonlinear and nonconstant coefficient problems that can be more efficiently implemented than Galerkin or tau approximations. Collocation is recommended for these problems. For example, the solution of the hyperbolic problem

$$\frac{\partial u}{\partial t} + e^{u+x}\frac{\partial u}{\partial x} = f(x,t) \quad (-1 \leq x \leq 1, \quad t > 0), \qquad (10.2)$$

$$u(-1,t) = 0,$$

would be difficult using Galerkin or tau approximation but is straightforward using collocation methods.

Let us explain how to march the solution to (10.2) forward by one time step efficiently using Chebyshev collocation. We introduce the $N+1$ collocation points $x_j = \cos \pi j/N$ $(j = 0,\ldots,N)$ and represent the current solution u_j as

$$u_j = \sum_{n=0}^{N} a_n \cos \frac{\pi n j}{N} . \tag{10.3}$$

Then we invert (10.3) by the fast Fourier transform to obtain a_n for $n = 0,1,\ldots,N$ and calculate

$$a_n^{(1)} = 2S_n/c_n$$

by (10.1). Next we evaluate

$$\left.\frac{\partial u}{\partial x}\right|_{x=x_j} = \sum_{n=0}^{N} a_n^{(1)} \cos \frac{\pi n j}{N} \tag{10.4}$$

using the fast Fourier transform. Finally, we evaluate $\exp(u_j+x_j)(\partial u/\partial x)_j$ at each of the 'grid' points x_j and use the results to march the solution forward to the next time step.

For quadratically nonlinear differential equations, like the Navier-Stokes equations of incompressible fluid dynamics, Galerkin and tau approximations are workable but normally require at least twice the computational work of collocation approximation. However, Galerkin approximation is sometimes very attractive because it gives approximations that are conservative and have no so-called aliasing errors (see Orszag 1971c, 1972 for a more complete discussion of these properties). Energy conservation properties of spectral methods are discussed at the end of Sec. 14.

Roundoff Errors

Transform methods normally give no appreciable amplification of roundoff errors. In fact, the evaluation of convolution-like sums using fast Fourier transforms often gives results with much smaller roundoff error than would be obtained if the convolution sums were evaluated directly.

On the other hand, the use of recurrence relations to evaluate derivatives can sometimes be a source of large roundoff errors. In this case, it is often best to convert the problem being solved into a new one that is numerically well-conditioned. An example of such a transformation is given below.

Example 10.1: Solution of $y''-ky=f(x)$ by Chebyshev polynomials

The boundary-value problem

$$y'' - ky = f(x) \qquad -1\le x\le 1 \tag{10.5}$$

$$y(-1) = A, \; y(1) = B$$

can be solved using a Chebyshev-tau approximation. The resulting approximation $y_N(x)$ is given by (see Appendix)

$$y_N(x) = \sum_{n=0}^{N} a_n \, T_n(x) \tag{10.6}$$

$$\frac{1}{c_n} \sum_{\substack{p=n+2 \\ p+n \text{ even}}}^{N} p(p^2-n^2)\, a_p - k\, a_n = f_n \quad (0 \le n \le N-2) \tag{10.7}$$

$$\sum_{n=0}^{N} (-1)^n a_n = A, \qquad \sum_{n=0}^{N} a_n = B, \tag{10.8}$$

where $\{f_n\}$ are the Chebyshev series coefficients of $f(x)$.

The solution of the system (10.7-8) for the Chebyshev coefficients $\{a_n\}$ may be done in several ways. One obvious way to do this efficiently is to write

$$a_n = a_n^{(1)} + \alpha a_n^{(2)} + \beta a_n^{(3)} . \tag{10.9}$$

Here $a_n^{(1)}$ satisfies $a_N^{(1)} = a_{N-1}^{(1)} = 0$ and

$$\frac{1}{c_n} \sum_{p=n+2}^{N} p(p^2-n^2) a_p^{(1)} - k a_n^{(1)} = f_n \quad (0 \le n \le N-2),$$

while $a_n^{(2)}$ satisfies $a_N^{(2)} = 1$, $a_{N-1}^{(2)} = 0$ and

$$\frac{1}{c_n} \sum_{p=n+2}^{N} p(p^2-n^2) a_p^{(2)} - k a_n^{(2)} = 0 \quad (0 \le n \le N-2),$$

and $a_n^{(3)}$ satisfies $a_N^{(3)} = 0$, $a_{N-1}^{(3)} = 1$, and

$$\frac{1}{c_n} \sum_{\substack{p=n+2 \\ p+n \text{ even}}}^{N} p(p^2-n^2) a_p^{(3)} - k a_n^{(3)} = 0 \quad (0 \le n \le N-2).$$

Each of the solutions $a_n^{(1)}$, $a_n^{(2)}$, $a_n^{(3)}$, may be found using roughly N operations by backwards recurrence. When the constants α and β in (10.9) are chosen so that the boundary conditions (10.8) are satisfied, a_n given by (10.9) satisfies (10.7-8).

The above procedure is efficient but it is not usually numerically stable. Roundoff errors multiply rapidly so that

a_n may have little significance.

A better procedure is to first convert (10.7-8) into a nearly tridiagonal system of equations. It may be shown that (10.7-8) is equivalent to the system

$$\frac{kc_{n-2}}{4n(n-1)} a_{n-2} - \left(1 + \frac{ke_{n+2}}{2(n^2-1)}\right) a_n + \frac{ke_{n+4}}{4n(n+1)} a_{n+2}$$

$$= \frac{c_{n-2} f_{n-2}}{4n(n-1)} - \frac{e_{n+2} f_n}{2(n^2-1)} + \frac{e_{n+4} f_{n+2}}{4n(n+1)} \quad (2 \leq n \leq N) \qquad (10.10)$$

with the boundary conditions (10.8) still applied. Here $c_0=2$, $c_n=1$ for $n>0$ and $e_n=1$ for $n \leq N$, $e_n=0$ for $n>N$. The system (10.8), (10.10) may be solved by standard banded matrix techniques in roughly the number of operations required to solve pentadiagonal systems of equations. The equations in the form (10.10) are essentially diagonally dominant so no appreciable accumulation of roundoff errors occurs. This technique for solution of (10.5) is very useful in implementing implicit spectral methods for dissipative terms and for solving Poisson-like equations (see Sec. 14).

11. Numerical Results for Hyperbolic Problems

We begin by presenting numerical results for spectral approximations to the problem

$$u_t + u_x = 0 \quad (-1\le x\le 1, t>0) \tag{11.1}$$

$$u(x,0) = 0\ , u(-1,t) = g(t), \tag{11.2}$$

whose exact solution is

$$u(x,t) = \begin{cases} g(t-x-1) & (x\le t-1) \\ 0 & (x>t-1). \end{cases} \tag{11.3}$$

If $g(t)$ is smooth, $u(x,t)$ is smooth for $|x|<1$ when $t>2$; when $t<2$, $u(x,t)$ is not smooth at $x=t-1$.

In Sec. 2 we explained how to obtain semi-discrete Galerkin, tau, and collocation approximation to (11.1-2) using either Chebyshev or Legendre polynomial expansions. In Sec. 9, we showed that either Adams-Bashforth or modified Euler time differencing gives stable and convergent results for these spectral approximations. The numerical results cited in this Section were obtained by Adams-Bashforth time-differencing; time steps were chosen small enough that time-differencing errors are negligible.

Comparison of Chebyshev and Legendre Polynomial Spectral Methods for Smooth Solutions

When $g(t) = -\sin M\pi t$, the solution (11.3) has M complete waves within $|x|\le 1$ when $t>2$. As argued in Sec. 3, we expect that accurate results will be obtained only if $N>M\pi$ polynomials are retained.

In Fig. 11.1, we plot the root-mean-square error for $|x| \leq 1$ averaged in time for $4 \leq t \leq 4.4$ obtained using the Chebyshev approximations to (11.1-2) when $g(t) = -\sin 5\pi t$. In this time interval, $u(x,t)$ is smooth for $|x| \leq 1$. Observe that the errors decrease exponentially fast when $N \gtrsim 5\pi$. Also observe that when the spectral approximations are relatively inaccurate (errors greater than roughly 10%), Galerkin approximation is most accurate followed by collocation and then tau. On the other hand, when the spectral approximations are very accurate (errors less than roughly 0.5%), tau approximation is most accurate followed by Galerkin and collocation. This behavior seems typical. Also observe from Fig. 11.1 that all three spectral approximations are nearly as accurate as the best (rms) Chebyshev approximation; in fact, tau approximation with $N+1$ polynomials is usually more accurate than the best approximation with N polynomials. Here the best (rms) Chebyshev approximation is that Nth degree polynomial that minimizes $\int_{-1}^{1} |u_N - u|^2 \, (1-x^2)^{-1/2} dx$.

In Fig. 11.2, we make similar comparisons of the error in spectral approximations using Legendre series for the problem (10.1-2) with $g(t) = -\sin 5\pi t$. Here too the errors decrease exponentially fast when $N \gtrsim 5\pi$. Again, tau approximation is more accurate than Galerkin when both are very accurate, while it is less accurate when both are relatively inaccurate. Also, tau approximation with $N+1$ polynomials and Galerkin approximations with $N+2$ polynomials are more accurate than the best Legendre approximation with N polynomials. Here the best Legendre approximation is that Nth degree polynomial that minimizes $\int_{-1}^{1} |u_N - u|^2 dx$.

Fig. 11.1. A plot of the L_2-errors in Chebyshev-spectral solution of (11.1-2) with $g(t) = -\sin 5\pi t$. The errors are averaged in time over the interval $4 \le t \le 4.4$; the exact solution $u(x,t) = \sin 5\pi(x+1-t)$ is smooth throughout this time interval. The best (rms) approximation is given by (3.41) with $M = 5$, $a = 1-t$ truncated after $T_N(x)$. Observe that the errors decrease rapidly for $N > 5\pi$.

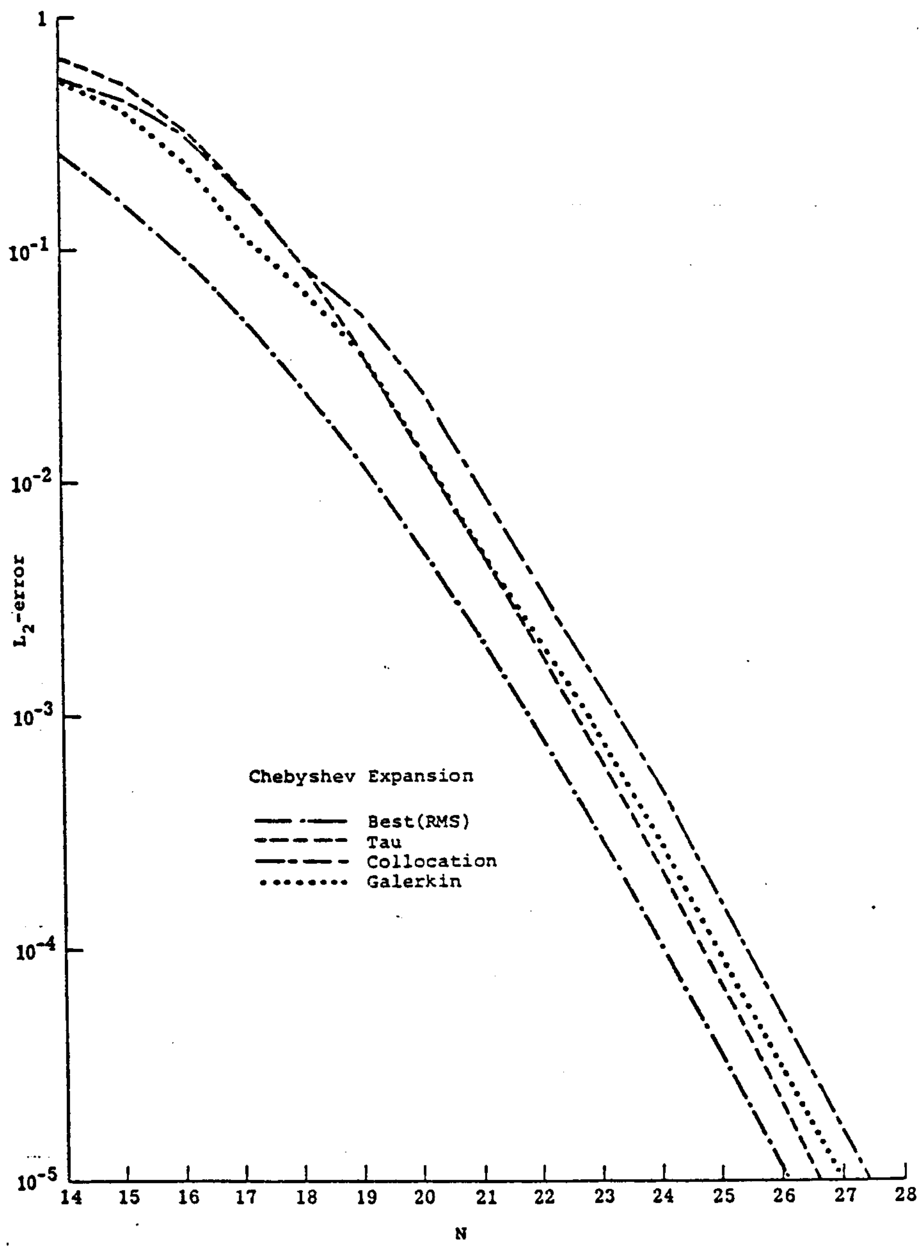
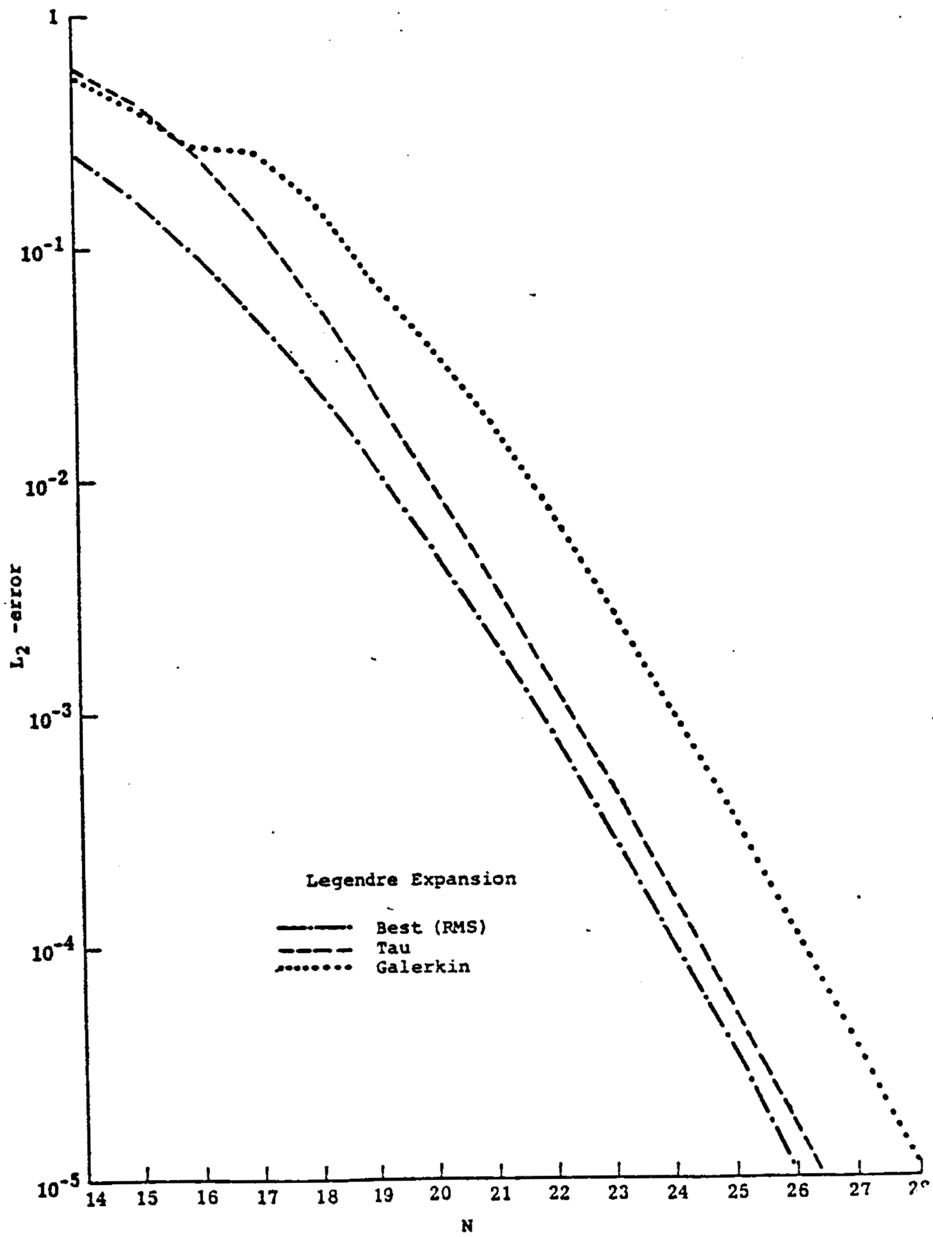


Fig. 11.2. Same as Fig. 11.1 except for Legendre-spectral solution of (11.1-2) with $g(t) = -\sin 5\pi t$. Here the best (rms) approximation is given by (3.45) with $M = 5$, $a = 1-t$ truncated after $P_N(x)$.



In Fig. 11.3-4 we plot the error $\varepsilon_N(x,t)$ in the best Chebyshev polynomial approximation to $\sin 5\pi(x+1-t)$ at $t=4$. Observe that $\varepsilon_N(x,t)$ is nearly an 'equal ripple' approximation (Acton 1970) so $u_N(x,t)$ is nearly a minimax approximation.

In Figs. 11.5-8 we plot the errors $\varepsilon_N(x,t)$ versus x at $t=4$ obtained by numerical solution of Chebyshev spectral approximations to (11.1-2). As N increases, the tau method gives the closest approximation to an equal-ripple error, which is consistent with the result shown in Fig. 11.1 that tau approximation gives the smallest errors at high accuracy.

In Figs. 11.9-10, we plot the error in the best Legendre polynomial approximation to $\sin 5\pi(x+1-t)$ at $t=4$. Observe that $\varepsilon_N(x,t)$ has large errors near the boundaries $x = \pm 1$. By comparing the results plotted in Figs. 11.3-4 with those plotted in Figs. 11.9-10, we conclude that the best Chebyshev polynomial approximation is closer to an equal ripple approximation than is the best Legendre polynomial approximation. Even though the best Legendre polynomial approximation to $u(x,t)$ gives the smallest mean-square error to u, the best Chebyshev polynomial approximation usually gives a smaller value of the maximum pointwise (L_∞) error. The large errors of the best Legendre approximation are concentrated near the boundaries $x=\pm 1$, while the Chebyshev weight function $(1-x^2)^{-1/2}$ tends to distribute the errors in the best Chebyshev approximation uniformly throughout $-1\le x\le 1$.

Fig. 11.3. A plot of the error $\epsilon_N(x,t)=u_N(x,t)-u(x,t)$ in the best (rms) Chebyshev polynomial approximation to $u(x,t)=\sin 5\pi(x+1-t)$ at $t=4$. Here $u_N(x,t)=\sum_{n=0}^{N} a_n(t)T_n(x)$ with $N = 20$ and $a_n(t)=(2/\pi c_n)\int_{-1}^{1} u(x,t)T_n(x)(1-x^2)^{1/2}dx$.

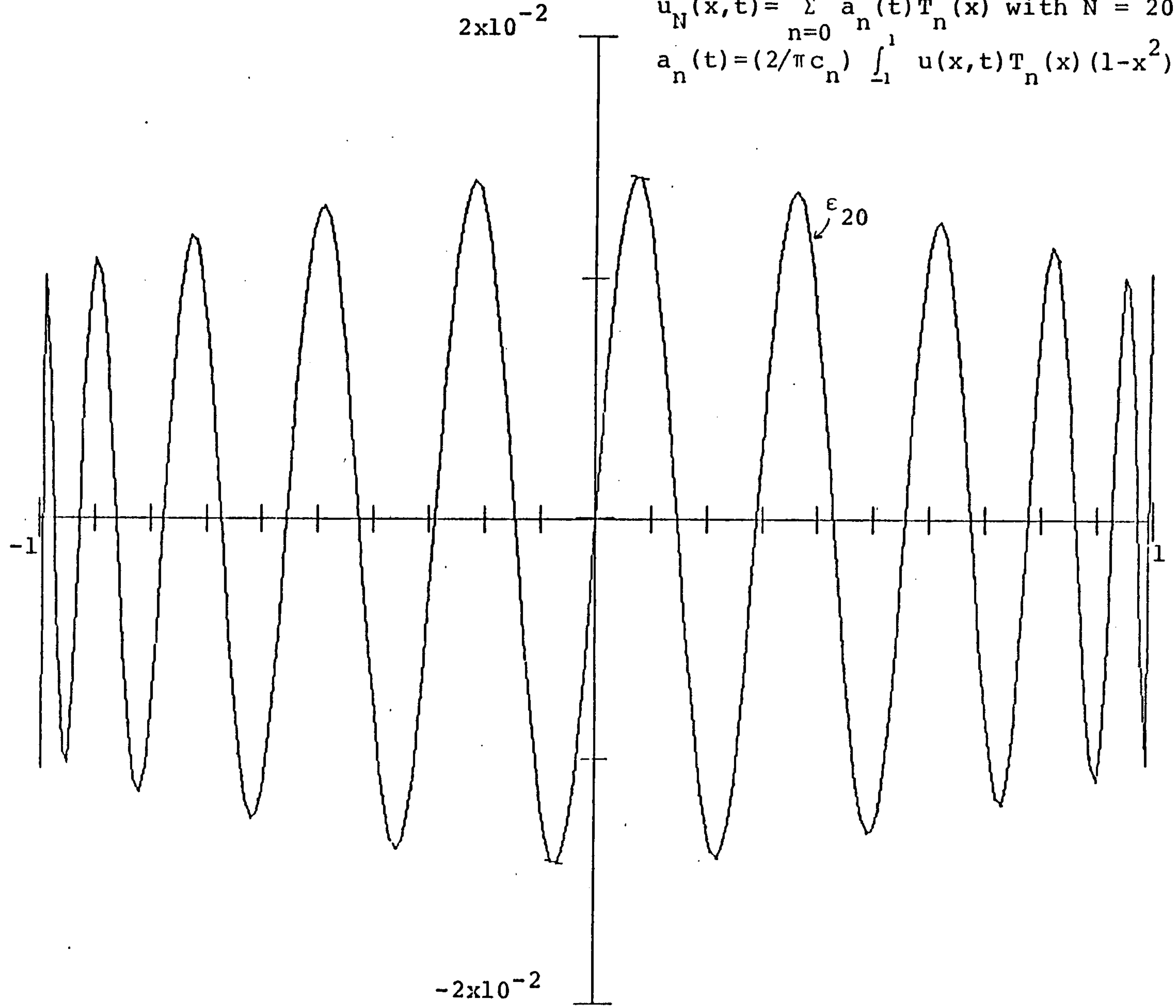
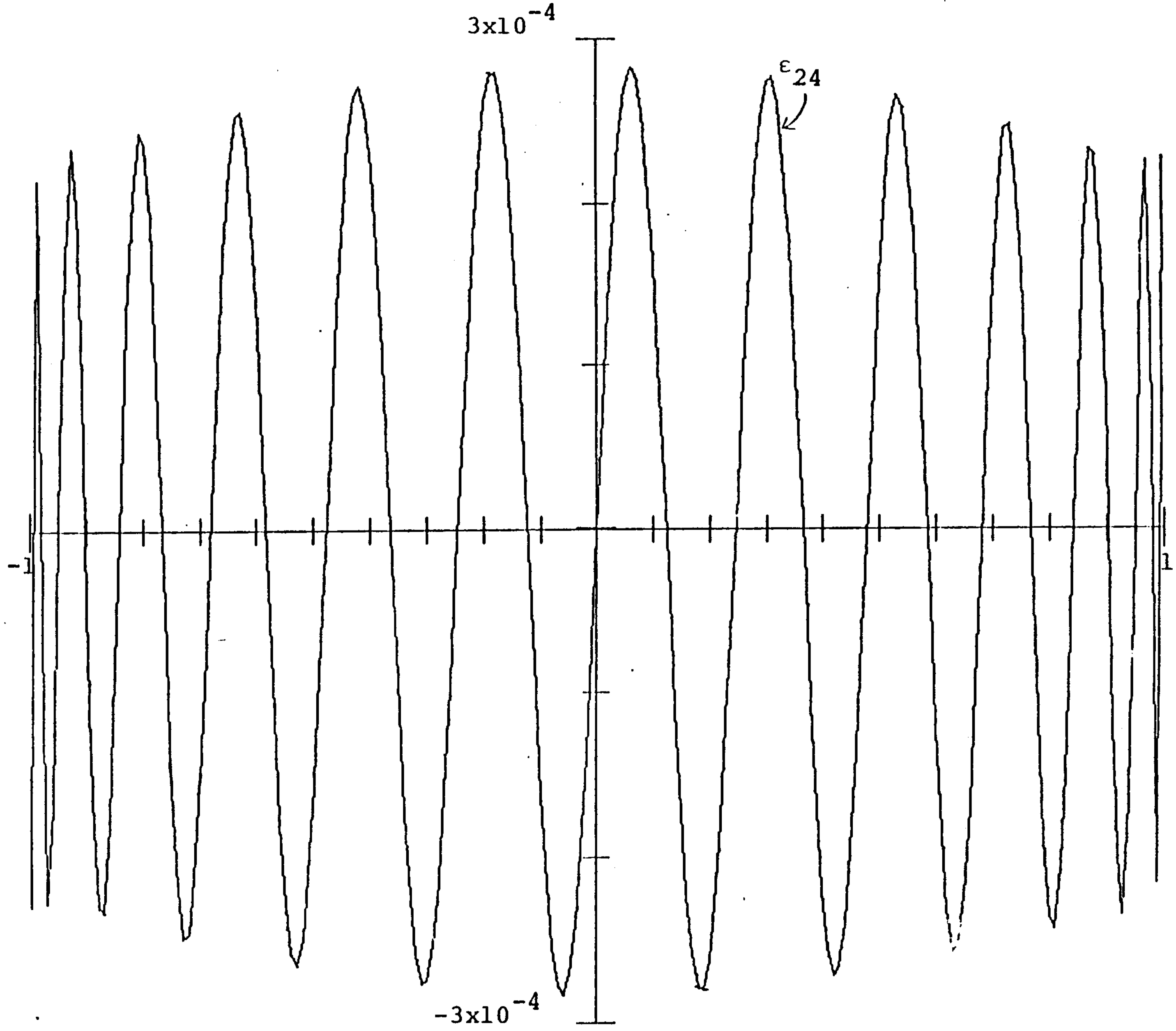


Fig. 11.4 Same as Fig. 11.3 except N=24.



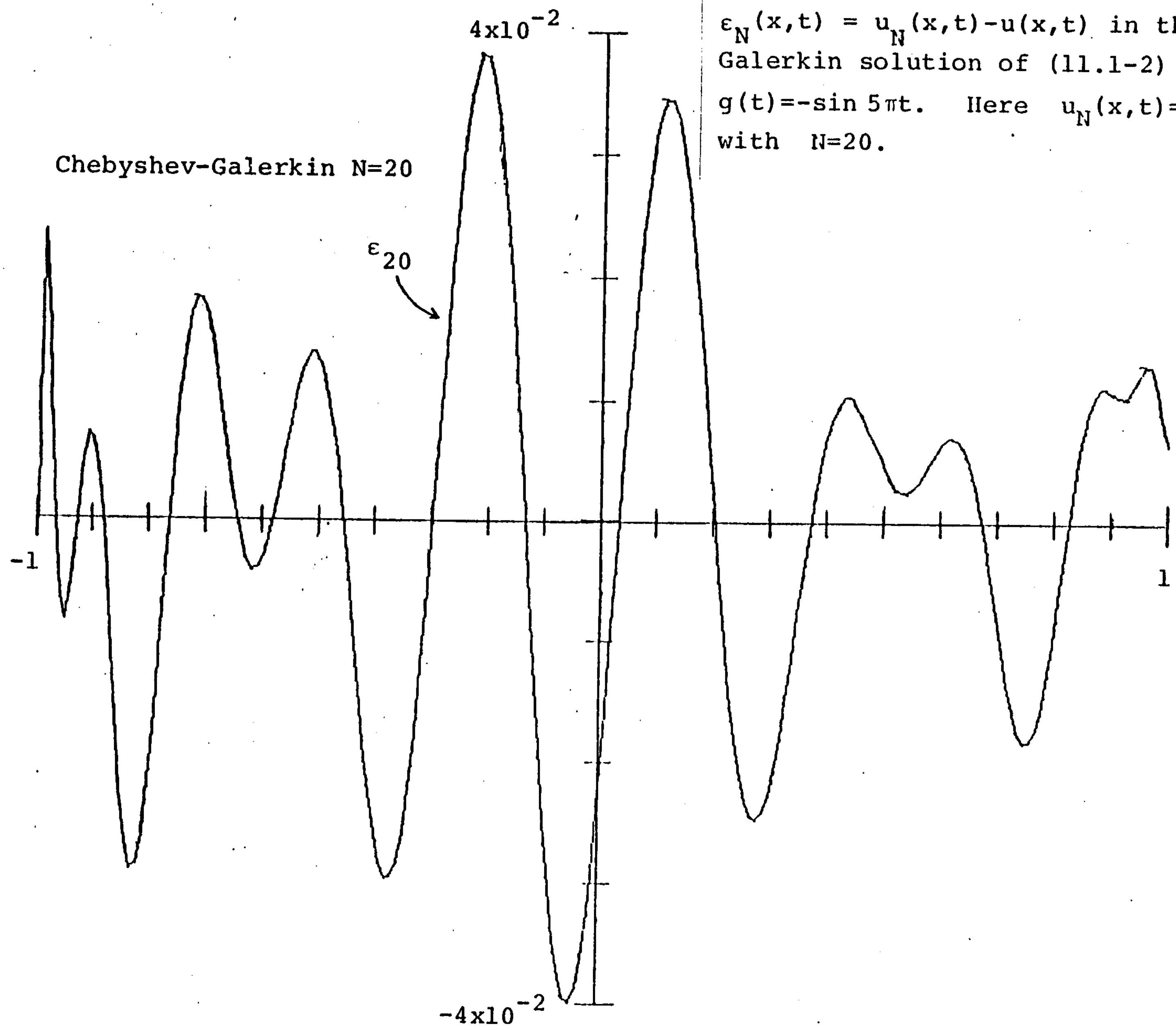


Fig. 11.5 A plot of the error $\varepsilon_N(x,t) = u_N(x,t) - u(x,t)$ in the Chebyshev-Galerkin solution of (11.1-2) at $t=4$ with $g(t) = -\sin 5\pi t$. Here $u_N(x,t) = \sum_{n=0}^{N} a_n(t) T_n(x)$ with $N=20$.

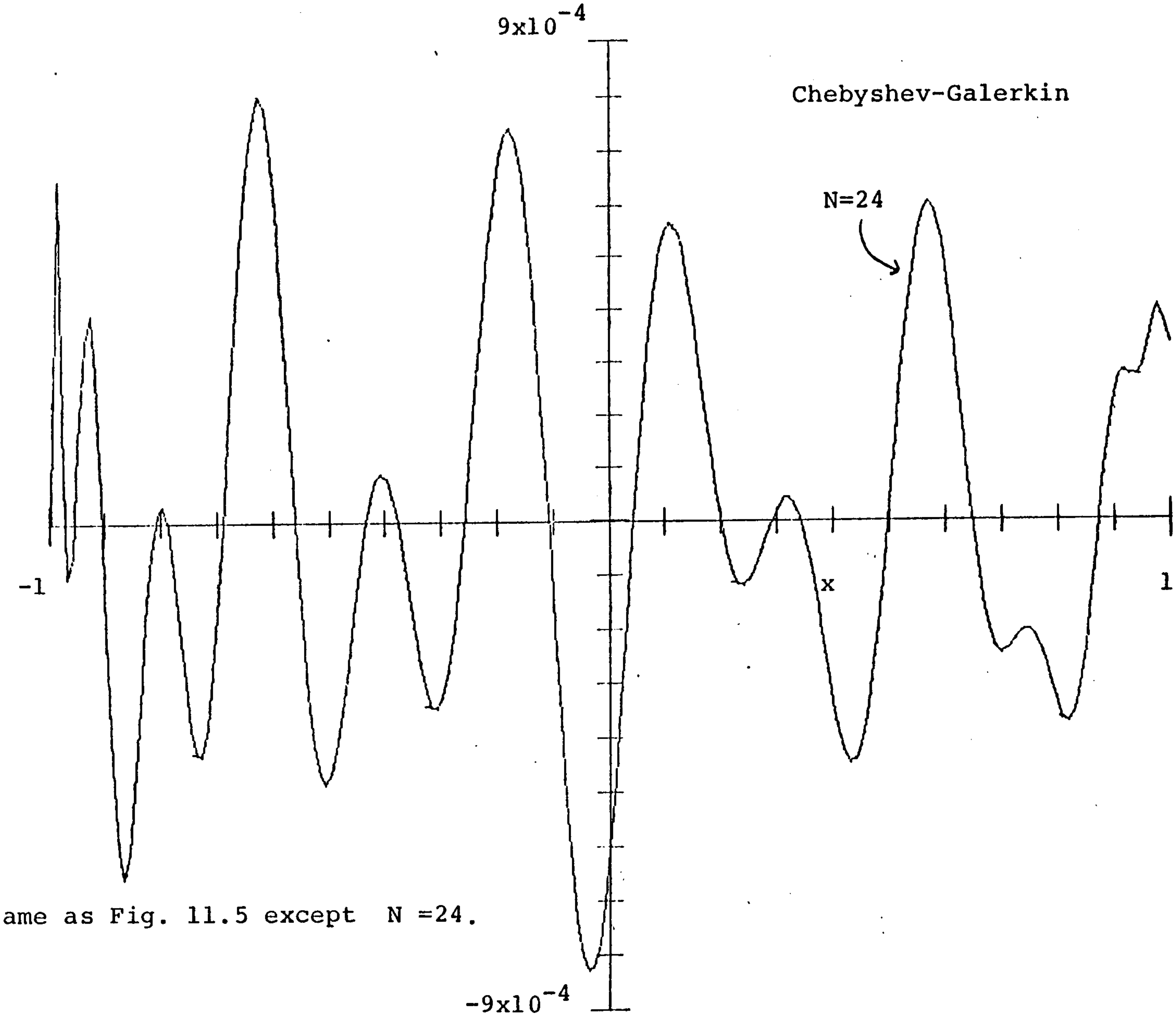
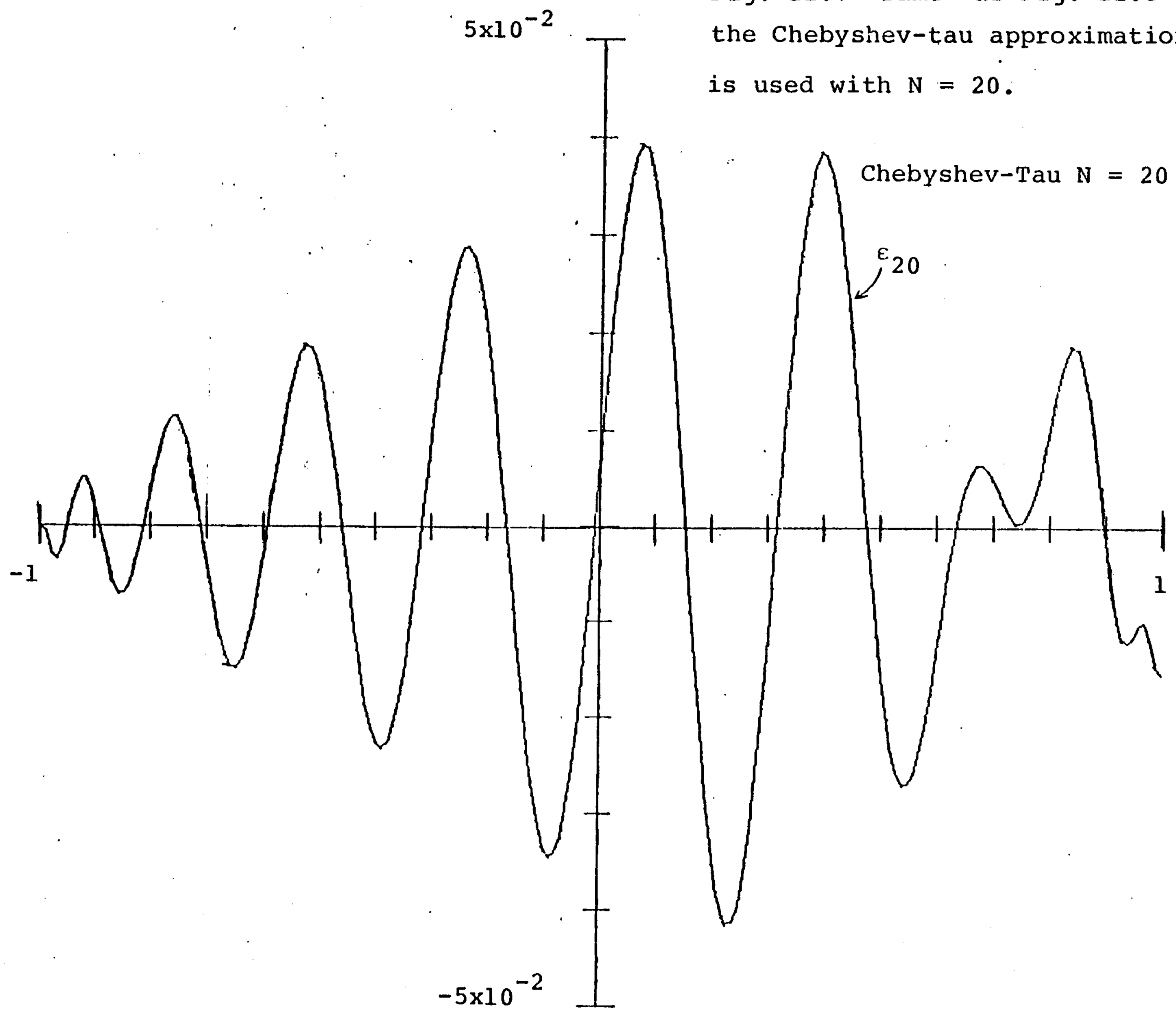


Fig. 11.6. Same as Fig. 11.5 except N =24.

Fig. 11.7 Same as Fig. 11.5 except that the Chebyshev-tau approximation to (11.1-2) is used with N = 20.



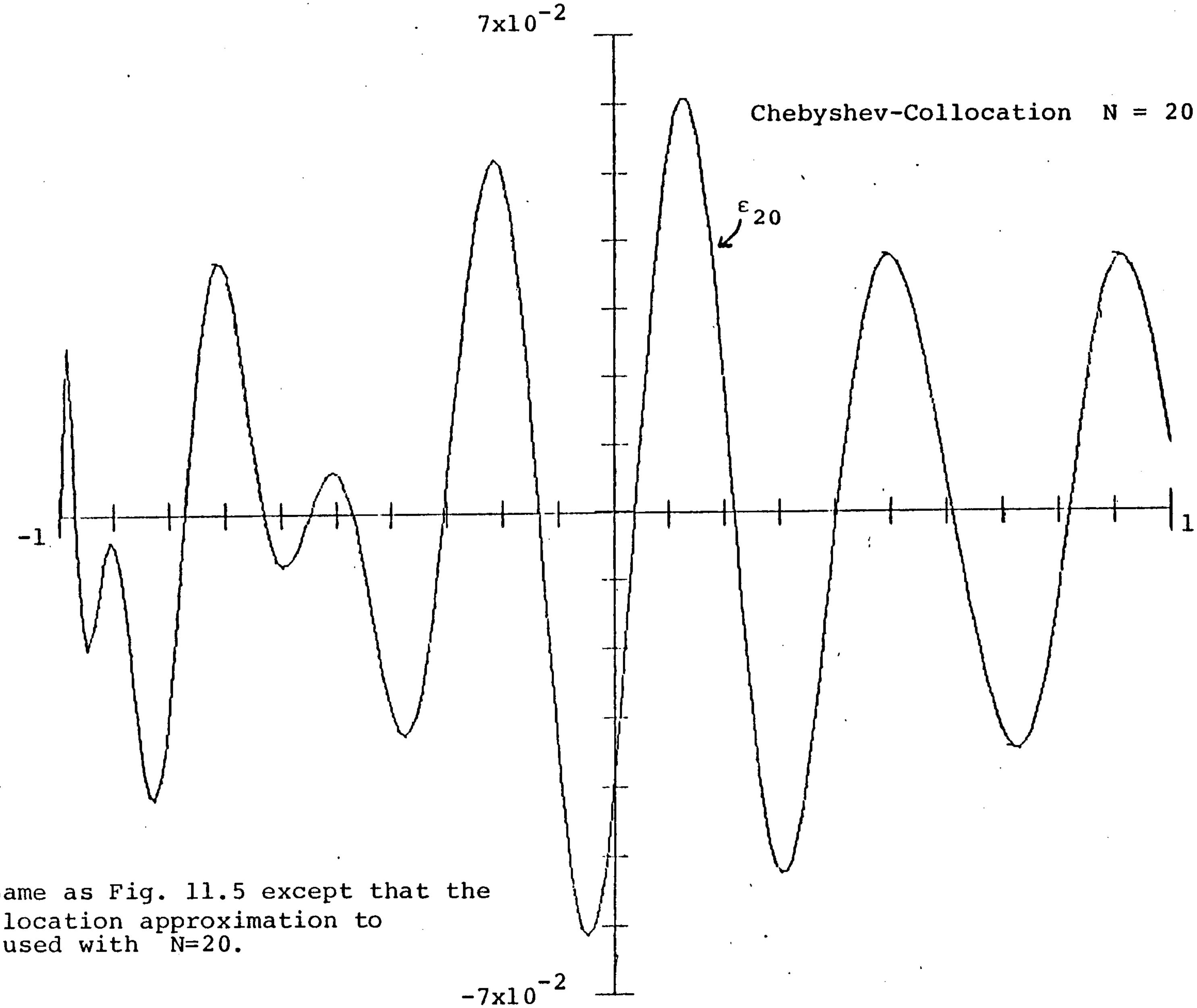


Fig. 11.8. Same as Fig. 11.5 except that the Chebyshev collocation approximation to (11.1-2) is used with N=20.

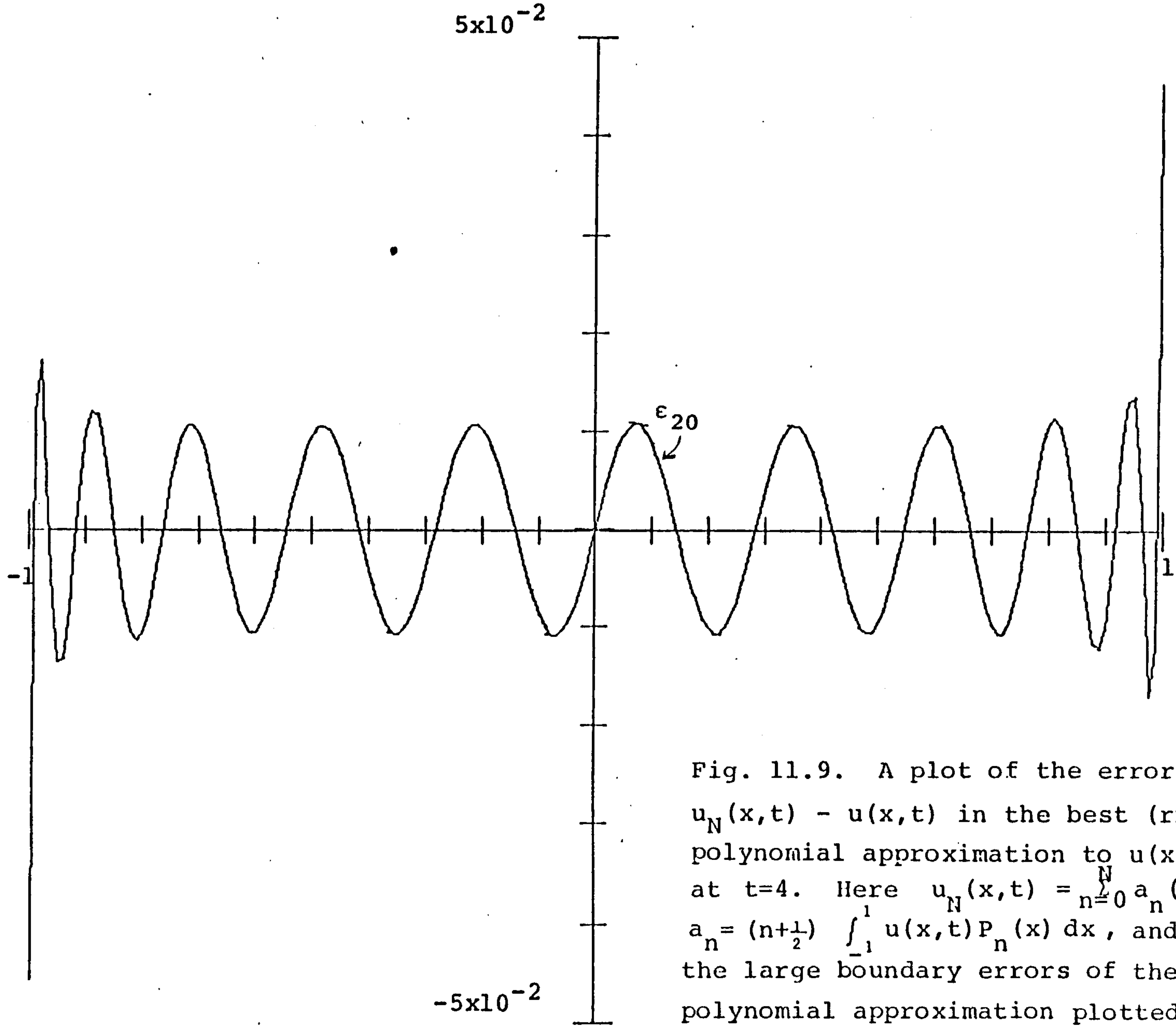
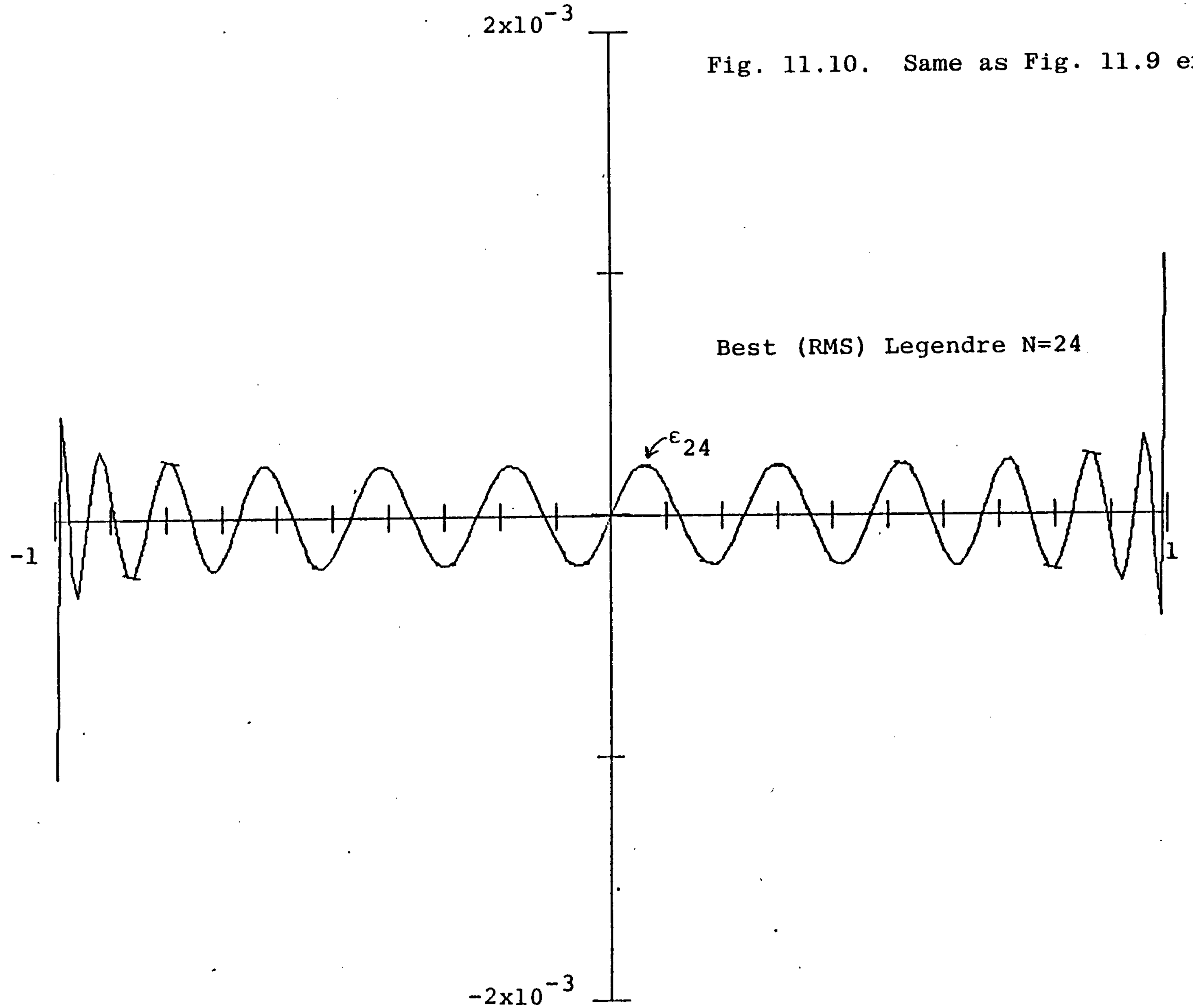


Fig. 11.9. A plot of the error $\epsilon_N(x,t) = u_N(x,t) - u(x,t)$ in the best (rms) Legendre polynomial approximation to $u(x,t) = \sin 5\pi(x+1-t)$ at $t=4$. Here $u_N(x,t) = \sum_{n=0}^{N} a_n(t) P_n(x)$, $a_n = (n+\frac{1}{2}) \int_{-1}^{1} u(x,t) P_n(x)\, dx$, and $N=20$. Observe the large boundary errors of the Legendre polynomial approximation plotted here in comparison with the best Chebyshev polynomial approximation plotted in Fig. 11.3.

Fig. 11.10. Same as Fig. 11.9 except N=24.



In Figs. 11.11-13, we plot the errors $\varepsilon_N(x,t)$ at $t=4$ obtained by numerical solution of Legendre spectral approximations to (11.1-2). As for Chebyshev-spectral approximations, the error in Legendre-tau approximation is smaller than that in Legendre-Galerkin approximation.

One important feature of Legendre-spectral approximation is that the spatial distribution of the error in tau and Galerkin approximation plotted in Figs. 11.11-13 differs markedly from the spatial distribution of the error in the best Legendre polynomial approximations plotted in Figs. 11.9-10. The boundary errors in the best L_2 approximation are relatively large while the boundary errors are relatively smaller in the spectral approximations.

The boundary (endpoint) errors in Legendre-tau approximation exhibit 'superconvergence' in the sense that they go to zero much faster than either the Legendre L_2 - errors or the L_2 and endpoint errors of Chebyshev-tau approximation. This fact is illustrated in Fig.11.14 where we plot the L_2 and endpoint errors of Legendre-tau and Chebyshev-tau spectral approximations to the solution of (11.1-2) with $g(t) = -\sin 5\pi t$. Here the endpoint error is $|u_N(+1,t)-u(+1,t)|$ at the outflow boundary $x= +1$.

Several features of the results plotted in Fig. 11.14 deserve comment. First, although the maximum error of the best N-term Chebyshev polynomial approximation is smaller than the maximum error of the best Legendre polynomial approximation to $u(x,t)$ by roughly a factor $1/\sqrt{N}$ [see (3.41) and (3.45)], the maximum error of the Legendre-tau approximation is smaller than the maximum error

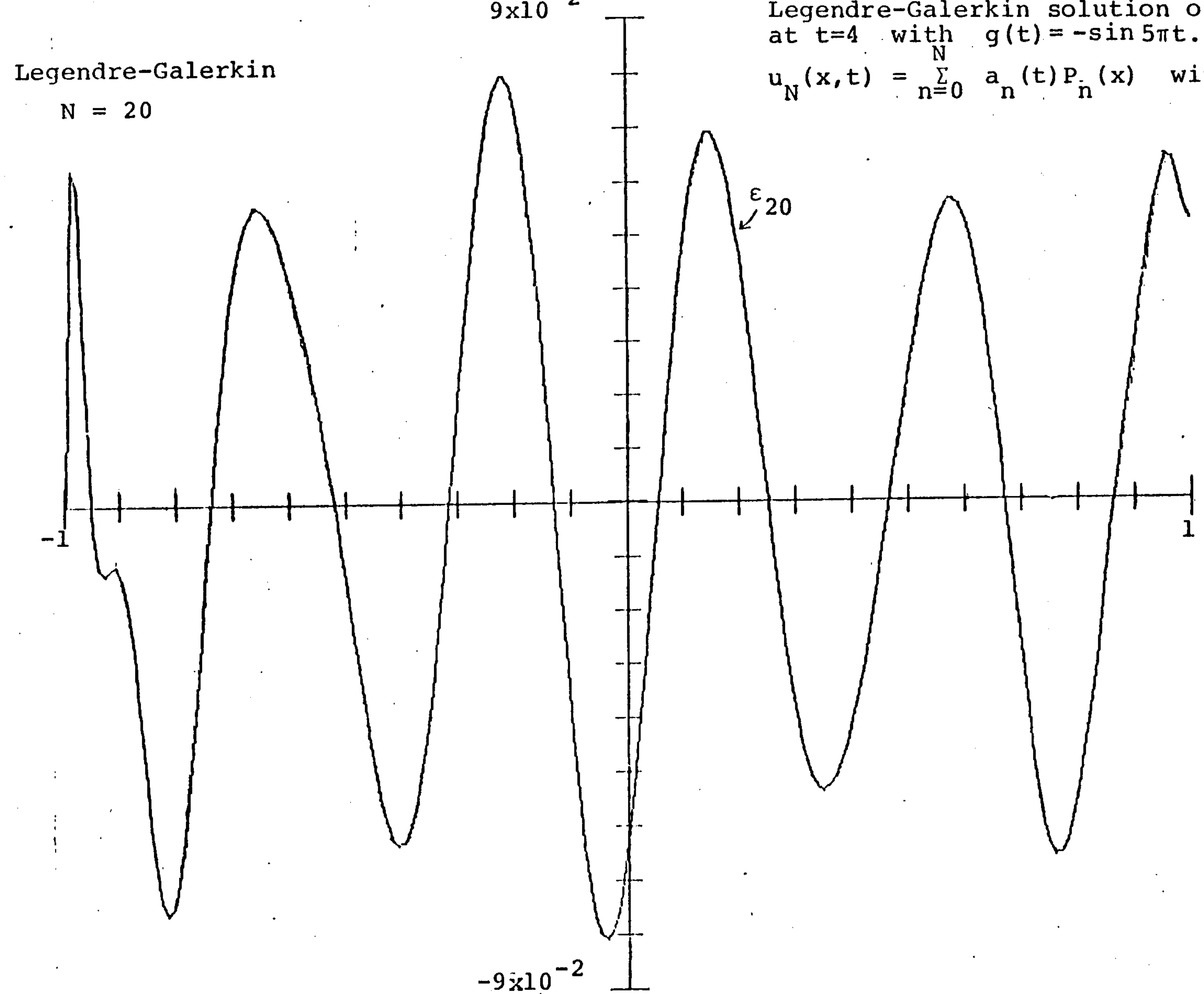


Fig. 11.11. A plot of the error $\varepsilon_N(x,t) = u_N(x,t) - u(x,t)$ in the Legendre-Galerkin solution of (11.1-2) at $t=4$ with $g(t) = -\sin 5\pi t$. Here $u_N(x,t) = \sum_{n=0}^{N} a_n(t) P_n(x)$ with $N = 20$.

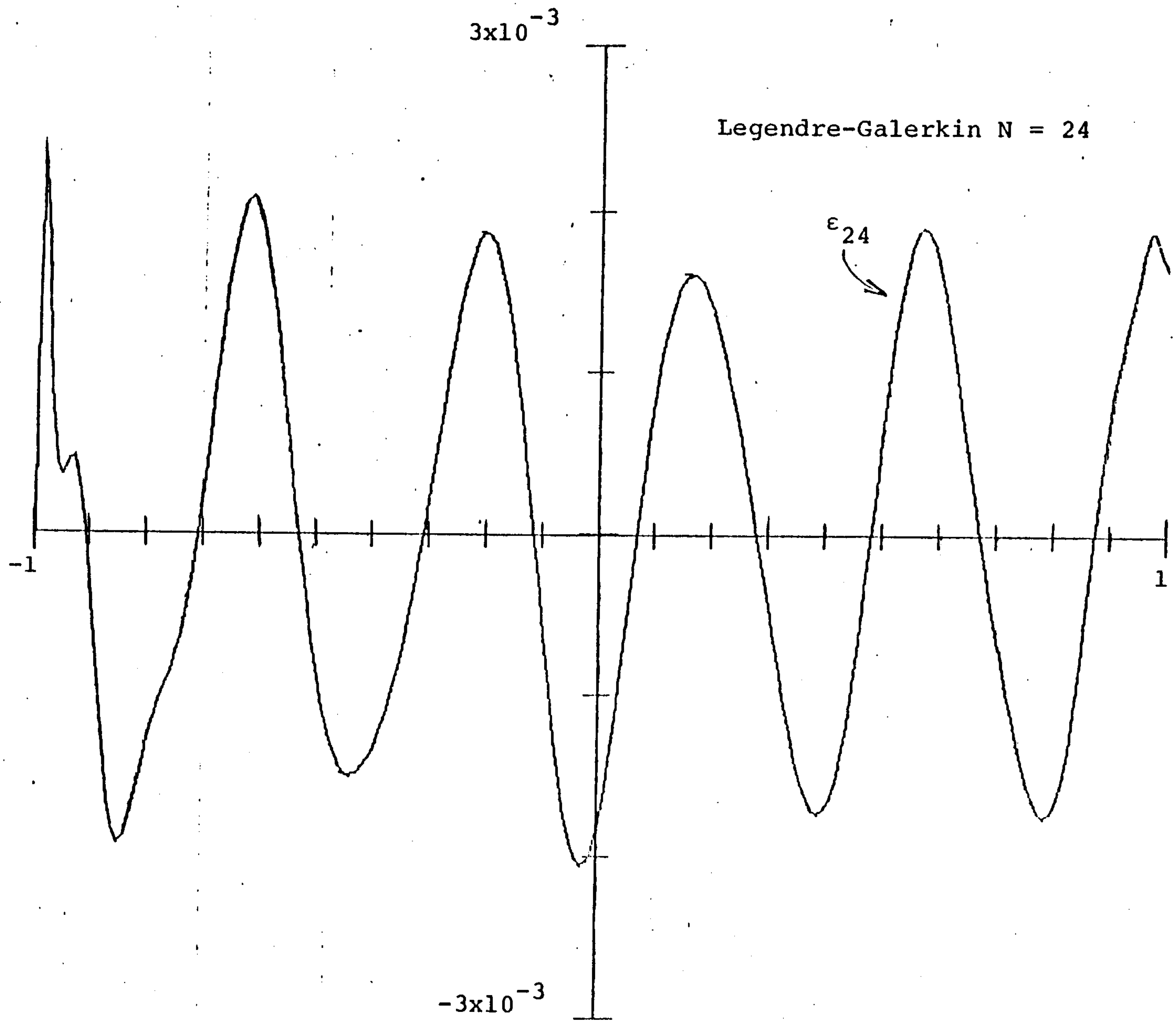


Fig. 11.12. Same as Fig. 11.11 except N=24.

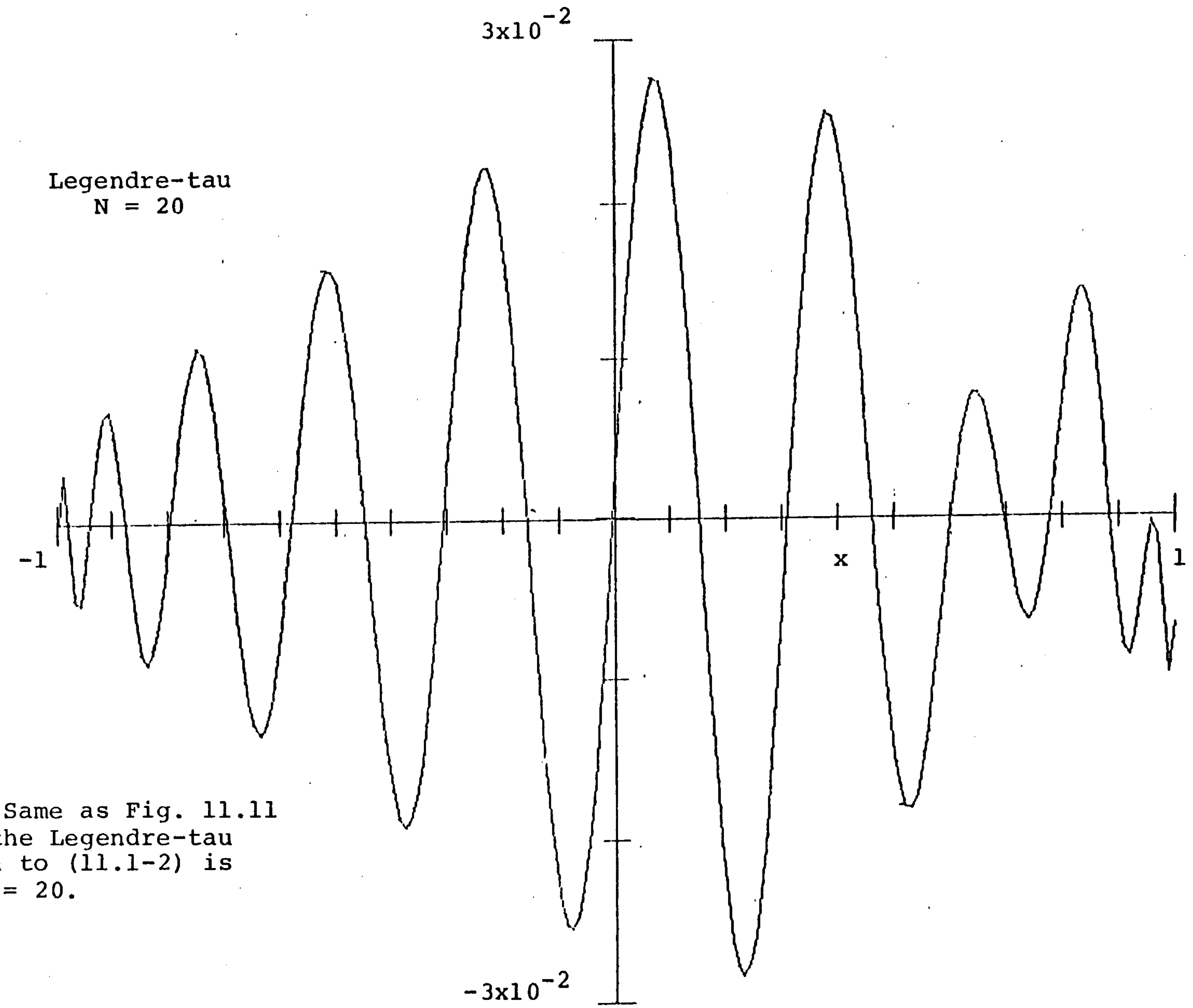
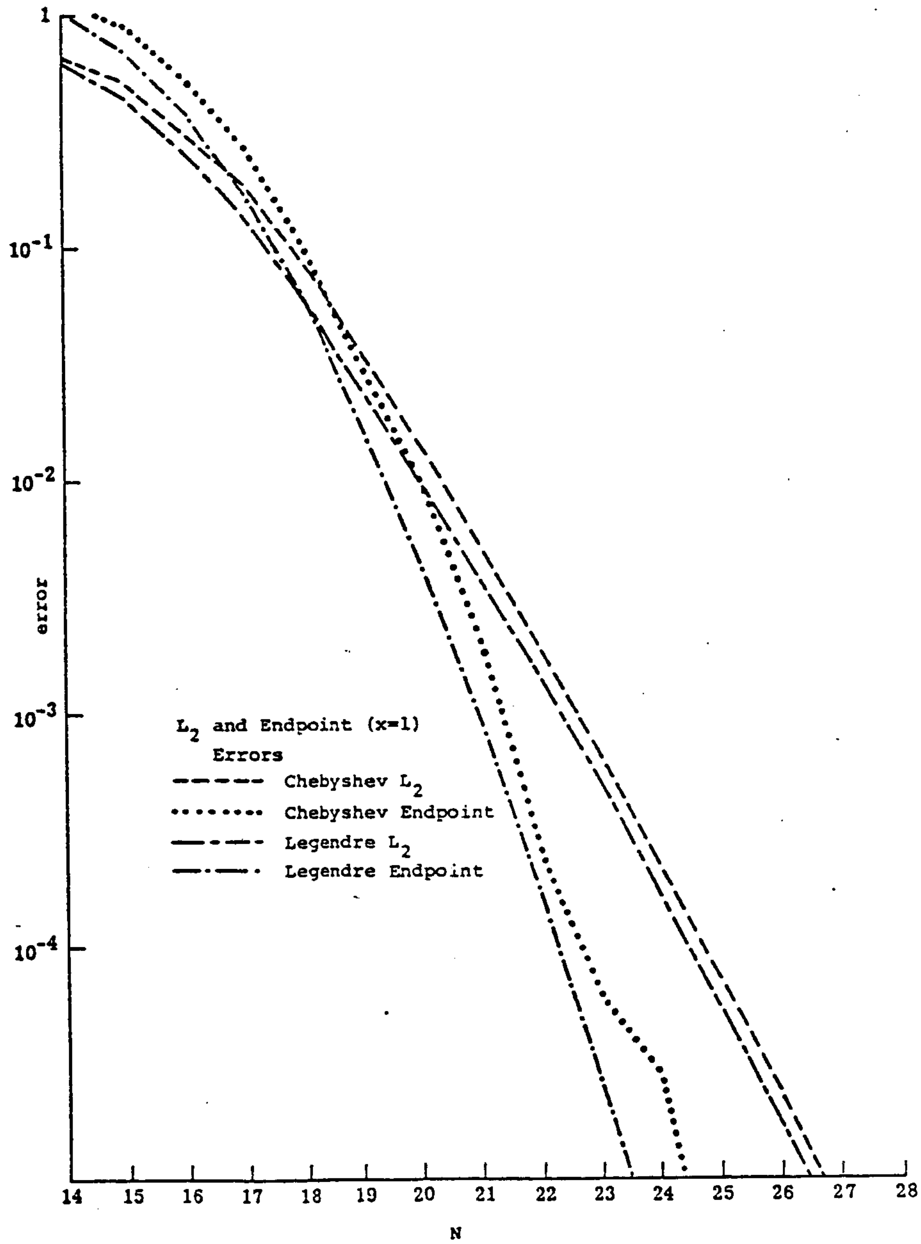


Fig. 11.13. Same as Fig. 11.11 except that the Legendre-tau approximation to (11.1-2) is used with $N = 20$.

Fig. 11.14 A comparison of the Chebyshev-tau and Legendre-tau L_2 and endpoint ($x = +1$) errors for the solution to (11.1-2) with $\tilde{g}(t) = -\sin 5\pi t$.



of the Chebyshev-tau approximation. Second, the endpoint error at $x=1$ of the Legendre-tau approximation goes to zero like the <u>square</u> of the endpoint error of the Chebyshev-tau approximation. This remarkable behavior of endpoint errors in Legendre-polynomial approximations was found originally by Lanczos in a slightly different context [Lanczos 1966 (p. 156), 1973].

A mathematical analysis of the errors of spectral approximations to (11.1-2) has been given recently by Dubiner (1977). Dubiner's results include: (a) asymptotic estimates of the errors incurred by the various spectral methods, including error oscillations when the solution is smooth; (b) a complete boundary layer description of the decay of large errors due to discontinuities after the discontinuities propagate out of the computational domain; (c) analysis of the behavior of the tau-function $\tau(t)$ in (2.34). Dubiner has analyzed a variety of spectral methods for (11.1-2) based on expansions in general Jacobi polynomials. His ingenious analyses of tau methods should permit more complete analysis of these methods than possible using earlier *a posteriori* analysis (see Fox & Parker 1968 for examples of *a posteriori* error analysis of tau methods).

Mesh Refinement

Sometimes it is useful to split up a domain into several subdomains and then use numerical methods of different spatial resolution in each. For example, in limited-area numerical weather forecasting near a metropolitan area, it may be desirable to have much finer resolution in a small region than is practical globally. One way to do this is to solve the problem separately on each of several subdomains and then to match the numerical solutions so obtained across subdomain boundaries. As a model of this procedure we consider the problem

$$u_t + u_x = 0 \qquad (-1\leq x\leq 1,\ t>0) \tag{11.4a}$$

$$u(-1,t) = g(t), \tag{11.4b}$$

$$v_t + v_x = 0 \qquad (1\leq x\leq 3, t>0) \tag{11.5a}$$

$$v(1+,t) = u(1-,t). \tag{11.5b}$$

With finite difference methods, the accurate solution of the coupled system (11.4.5) using different grids for $-1\leq x\leq 1$ than for $1\leq x\leq 3$ may be troublesome. Inaccurate results or even numerical instabilities can result from the matching (Browning, Kreiss & Oliger 1973). Because grids with different grid separations have different dispersion characteristics for waves propagating on the grid, waves can reflect from the boundary at $x=1$ and cause large errors.

Spectral methods are attractive for the solution of mesh refinement problems like (11.4-5) because they give small endpoint errors. For example, the Chebyshev-tau approximation to (11.4-5) with $N+1$ polynomials to represent the solution for $-1\leq x\leq 1$ and $M+1$ polynomials to represent the solution for $1\leq x\leq 3$ is given by

$$u_N(x,t) = \sum_{n=0}^{N} a_n(t)\ T_n(x) \qquad (-1\leq x\leq 1) \tag{11.6}$$

$$v_M(x,t) = \sum_{m=0}^{M} b_m(t)\ T_m(x-2) \qquad (1\le x\le 3) \tag{11.7}$$

$$\frac{da_n}{dt} = -\frac{2}{c_n} \sum_{\substack{p=n+1 \\ p+n \text{ odd}}}^{N} p\ a_p \qquad (0\le n\le N-1) \tag{11.8}$$

$$\frac{db_m}{dt} = -\frac{2}{c_m} \sum_{\substack{p=m+1 \\ p+m \text{ odd}}}^{M} p\ b_p \qquad (0\le m\le M-1) \tag{11.9}$$

$$\sum_{n=0}^{N} (-1)^n\ a_n = g(t) \tag{11.10}$$

$$\sum_{m=0}^{M} (-1)^m\ b_m = \sum_{n=0}^{N} a_n \tag{11.11}$$

It may easily be shown that if $g(t)$ is smooth, the solution to (11.6-11) converges to the solution of (11.4-5) throughout $-1\le x\le 3$ faster than any finite power $1/N$ or $1/M$ as $N, M\to\infty$. The solutions for $-1\le x\le 1$ and $1\le x\le 3$ match without the necessity of imposing any matching conditions in addition to (11.5b) or (11.11). Because no spurious downstream boundary conditions are applied at $x=+1$ on the wave propagating in the interval $-1\le x\le 1$, there are no reflected waves.

One more example of a refined mesh spectral calculation is instructive. Consider the heat equation problem

$$\frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial x^2} \qquad -1 \le x \le 1 \tag{11.12a}$$

$$\frac{\partial v}{\partial t} = \nu \frac{\partial^2 v}{\partial x^2} \qquad 1 \le x \le 3 \tag{11.12b}$$

$$u(-1,t) = f(t), \quad v(3,t) = g(t) \tag{11.12c}$$

$$u(1-,t) = v(1+,t), \quad \frac{\partial u}{\partial x}(1-,t) = \frac{\partial v}{\partial x}(1+,t) \tag{11.12d}$$

where (11.12d) follows by requiring continuity of temperature and heat flux across the boundary at $x=1$. A Chebyshev-tau approximation to (11.12) is given by (11.6-7) with

$$\frac{da_n}{dt} = \frac{\nu}{c_n} \sum_{\substack{p=n+2 \\ p+n \text{ even}}}^{N} p(p^2-n^2)\, a_p \qquad (0 \le n \le N-2) \tag{11.13a}$$

$$\frac{db_m}{dt} = \frac{\nu}{c_m} \sum_{\substack{p=m+2 \\ p+m \text{ odd}}}^{M} p(p^2-m^2)\, b_p \qquad (0 \le p \le M-2) \tag{11.13b}$$

$$\sum_{n=0}^{N} (-1)^n a_n = f(t), \quad \sum_{m=0}^{M} b_m = g(t) \tag{11.13c}$$

$$\sum_{n=0}^{N} a_n = \sum_{m=0}^{M} (-1)^m b_m, \quad \sum_{n=0}^{N} n^2 a_n = - \sum_{m=0}^{M} (-1)^m m^2 b_m . \tag{11.13d}$$

It may be shown as in Example 7.1(v) that this approximation is semi-bounded and hence stable and convergent.

Discontinuities

When $t<2$, the solution (11.3) to (11.1-2) is not smooth at $x=t-1$; if $g(t) = \sin M\pi t$, the solution has a discontinuous derivative. This discontinuity seriously degrades the rate of convergence of spectral approximations near the discontinuity. Nevertheless, spectral approximations are still normally much more accurate than finite-difference approximations to the same problem. Orszag & Jayne (1974) give comparisons between finite-difference and spectral approximations to discontinuous solutions; in particular, they argue that if the pth derivative of the solution is discontinuous, the rate of convergence of Chebyshev-spectral approximations to (11.1-3) for $t<2$ is of order $1/N^p$ as $N\to\infty$. Dubiner (1977) has given a detailed asymptotic analysis of this problem. His results include detailed behavior of the error for all x and t and are in good agreement with numerical solutions.

One of the attractive features of spectral methods for problems with discontinuities is that the region of large errors is more localized near the discontinuity than in finite-difference methods. Thus, it should be possible to eliminate oscillations near the discontinuity using less dissipation than is required when finite difference methods are used. A comparison of the error in Chebyshev-tau and second and fourth-order finite-difference solutions of (11.1-2) for $t<2$ is given in Fig. 11.15.

Another interesting way to use spectral methods for problems with discontinuous solutions has been suggested by Boris & Book (1976). The "optimal flux-corrected transport" approximation gives good resolution of discontinuities without introduction of unphysical numerical oscillations near the discontinuity. The idea is to add in an artificial diffusion to smooth the discontinuity and then to 'anti-diffuse' the resulting solution in such a way that no new oscillations or maxima/minima are produced.

Comparison with Finite Difference Methods

Finite-difference approximations to (11.1-2) must be formulated carefully near the boundaries $x = \pm 1$. For example, the fourth-order semi-discrete approximation

$$\frac{\partial u_j}{\partial t} + \frac{8(u_{j+1}-u_{j-1})-u_{j+2}+u_{j-2}}{12\Delta x} = 0$$

where $u_j(t) = u(j\Delta x,t)$, must be modified at $x=-1+\Delta x,\ 1-\Delta x,1$ because $u(-1-\Delta x,t)$, $u(1+\Delta x,t)$, $u(1+2\Delta x,t)$ all lie outside the

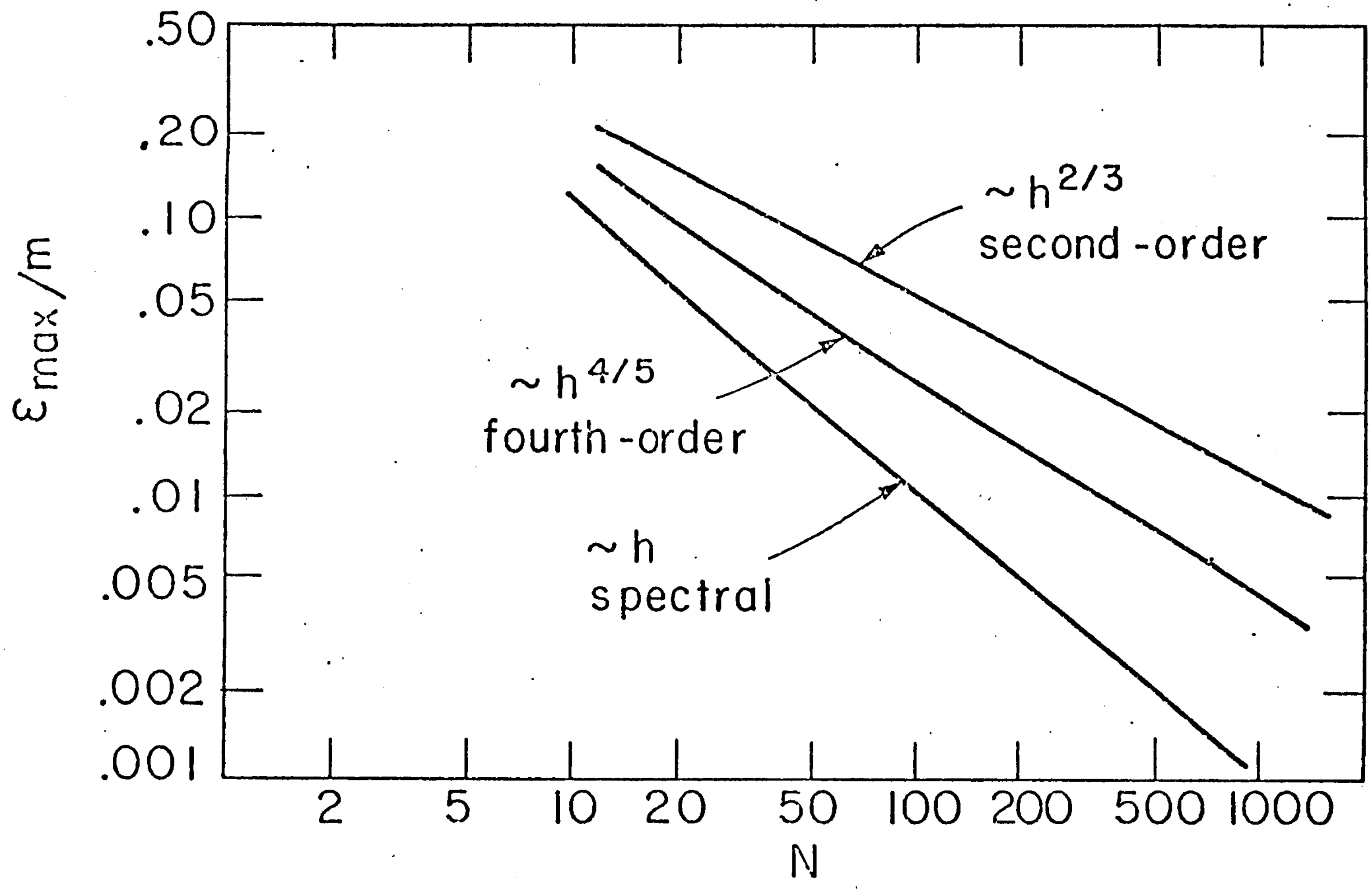


Fig. 11.15 Maximum pointwise errors in the solution of (11.1-2) with $g(t) = \sin m\pi t$ at $t = 1.2$ when the discontinuity in the solution (11.3) is at $x = 0.2$. Here ϵ_{max} is the maximum error and $N = 2/h$ is the number of grid points, or of Chebyshev polynomials in the spectral method.

computational domain $-1 \leq x \leq 1$. Kreiss & Oliger (1973) discuss methods to formulate difference approximations at these grid points. However, it is not known how to formulate appropriate 'boundary' conditions for arbitrary order difference schemes. This difficulty is an artifact of difference methods; a fourth-order difference quation requires 4'boundary' conditions while only 1 condition (11.2) is properly imposed on the first-order differential equation (11.1).

On the other hand, properly formulated spectral methods require no 'spurious' boundary conditions. Indeed, the imposition of a spurious boundary condition on a spectral approximation to (11.1), like $\partial u/\partial x = 0$ at $x = +1$, will induce an unconditional instability (see Sects. 8,12). The mathematics of spectral approximations follows closely the mathematics of the differential equation being solved.

Spectral approximations also require considerably fewer degrees of freedom to achieve accurate results than are required by difference methods. A comparison for the problem (11.1-2) is given in Table 11.1 for late times at which the solution is smooth.

In Figs. 11.16-19 we show three-dimensional perspective plots of the solution to a simple two-dimensional hyperbolic problem with periodic boundary conditions

$$\frac{\partial A(x,y,t)}{\partial t} - y\,\frac{\partial A(x,y,t)}{\partial x} + x\,\frac{\partial A(x,y,t)}{\partial y} = 0 \qquad (11.14)$$

Table 11.1

Second-order			Fourth-order			Chebyshev-tau		
N	M	ε_2	N	M	ε_4	N	M	ε_∞
40	2	0.1	20	2	0.04	16	4	0.08
80	2	0.03	30	2	0.008	20	4	0.001
160	2	0.008	40	2	0.002	28	8	0.2
40	4	1.	40	4	0.07	32	8	0.008
80	4	0.2	80	4	0.005	42	12	0.2
160	4	0.06	160	4	0.0003	46	12	0.02

Table 11.1. L_2 (rms) errors for the solution of (11.1-2) with $g(t) = \sin M\pi t$. The errors listed are measured at $t=5$ when the solution (11.3) is smooth. Time differencing errors are negligible and N is the number of grid points or Chebyshev polynomials. Observe that to achieve a 1% error, the second-order method requires $N/M \gtrsim 40$, the fourth-order method requires $N/M \gtrsim 15$, while the Chebyshev-tau method requires $N/M \gtrsim \pi$.

with

$$A(x \pm 2\pi, y \pm 2\pi, t) = A(x,y,t).$$

The solution to (11.14) is constant along the characteristics $x+iy = (x_0 + iy_0)e^{it}$. Therefore, $A(x,y,2\pi) \equiv A(x,y,0)$ so the solution should keep A unchanged after a time 2π. In Fig. 11.16, we plot the initial conditions used for the calculation whose results are plotted in Figs. 11.17-19. In Fig. 11.17 we plot the results at $t=2\pi$ of a second-order centered space difference scheme; in Fig. 11.18 we plot the results of a fourth-order scheme and in Fig. 11.19 we plot the results of a spectral calculation using the Fourier expansion

$$A(x,y,t) = \sum_{|k| \le K} \sum_{|p| \le P} a(k,p,t) e^{ikx+ipy} .$$

All three calculations used the same number of degrees of freedom but the differences in accuracy are striking. The Fourier-spectral method works well even though the convecting velocity $(-y,x)$ in (11.14) has jump discontinuities at $x = \pm 2\pi$, $y = \pm 2\pi$. The dominant error in all three calculations originates from the 'corners' of the cone in the initial $A(x,y,0)$ distribution; this error appears as a large lagging phase error in the finite difference solutions which explains the 'wakes' of large errors following the remnants of $A(x,y,2\pi)$.

<u>Higher-Order Wave Equations</u>

The mixed initial-boundary value

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} \quad (-1 \le x \le 1, t > 0) \tag{11.15}$$

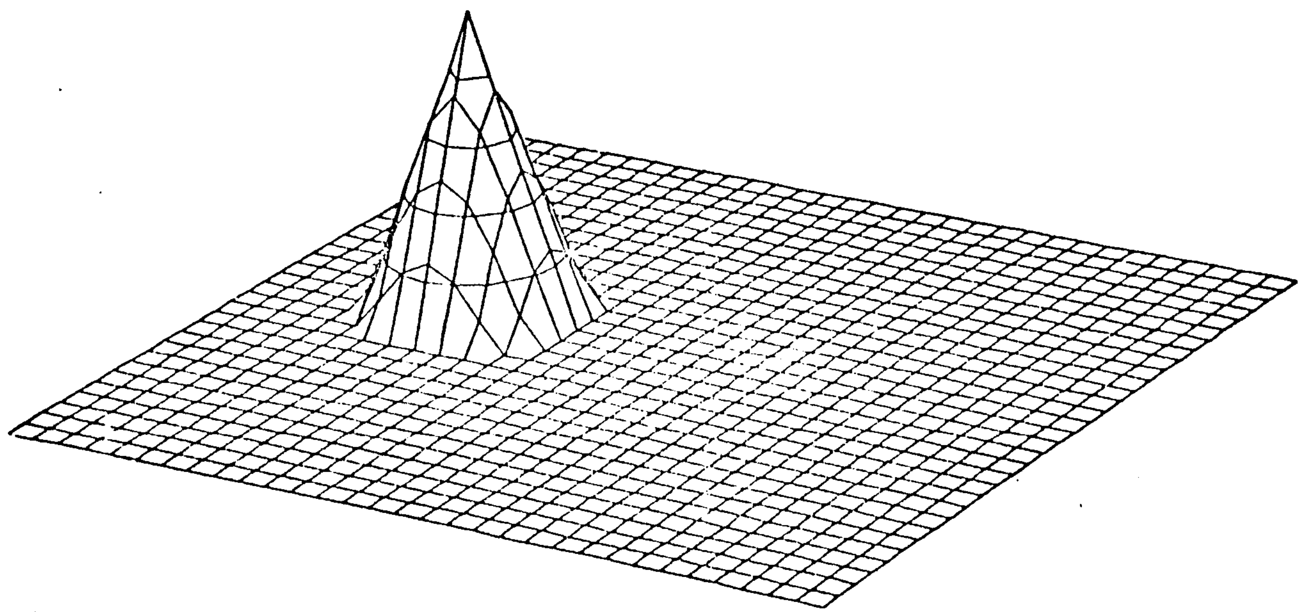


Fig. 11.16 A perspective plot of the $A(x,y,t=0)$ used in a numerical test of methods to solve (11.14).

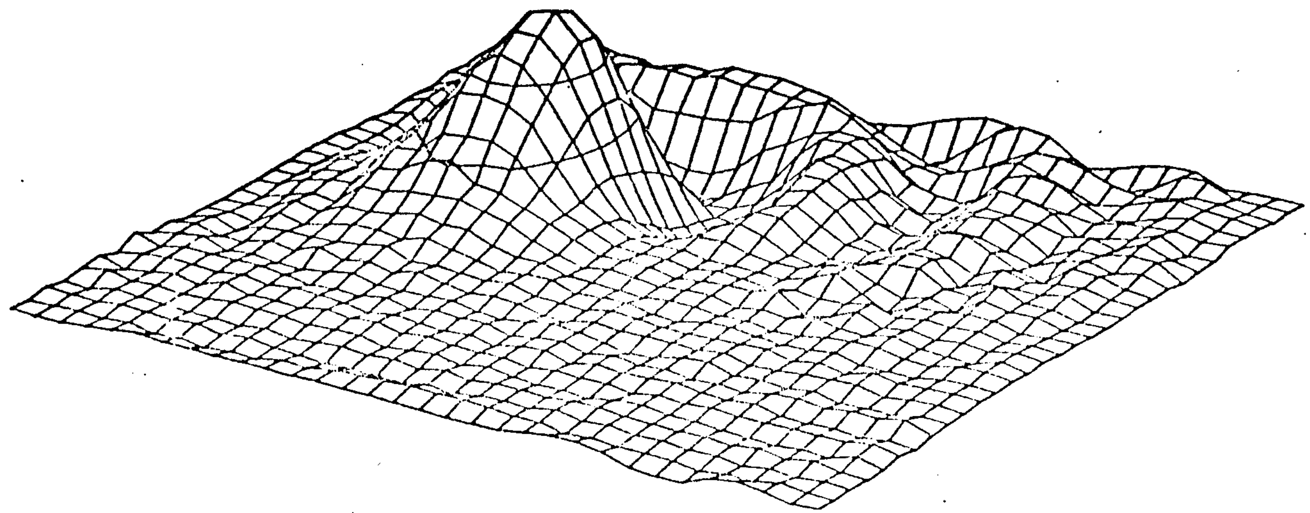


Fig. 11.17 Three-dimensional perspective plot of the $A(x,y,t)$ field obtained after one revolution using a second-order centered difference scheme on a 32 x 32 grid.

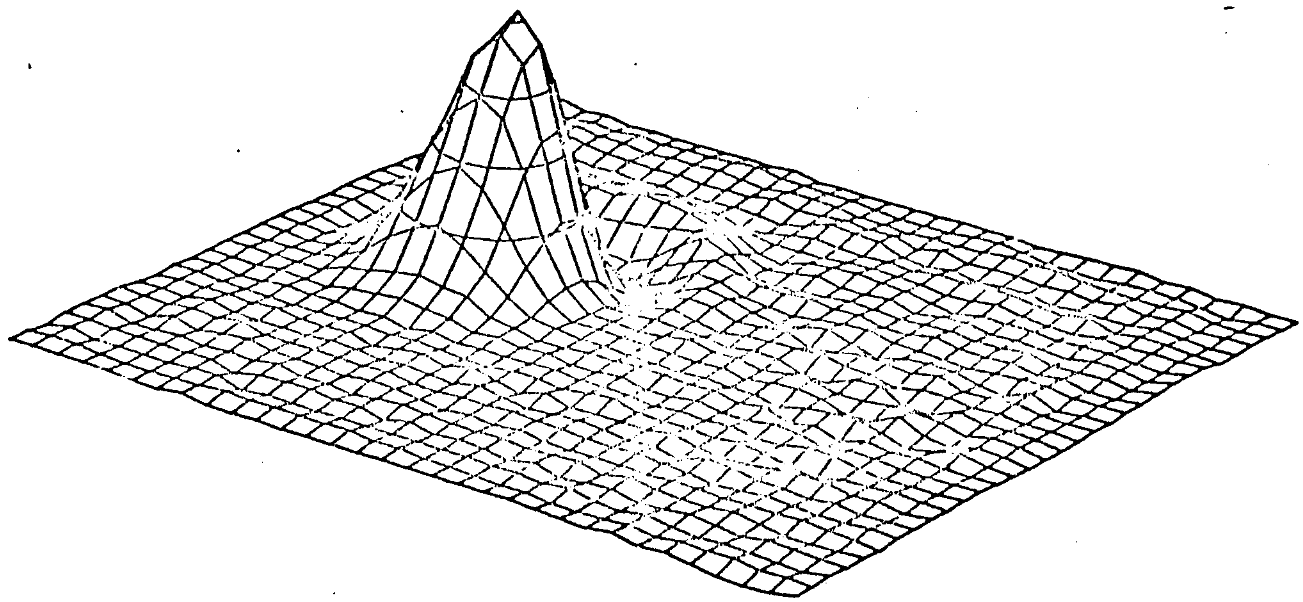


Fig. 11.18 Same as Fig. 11.17 except using a fourth-order centered difference scheme on a 64 x 64 grid.

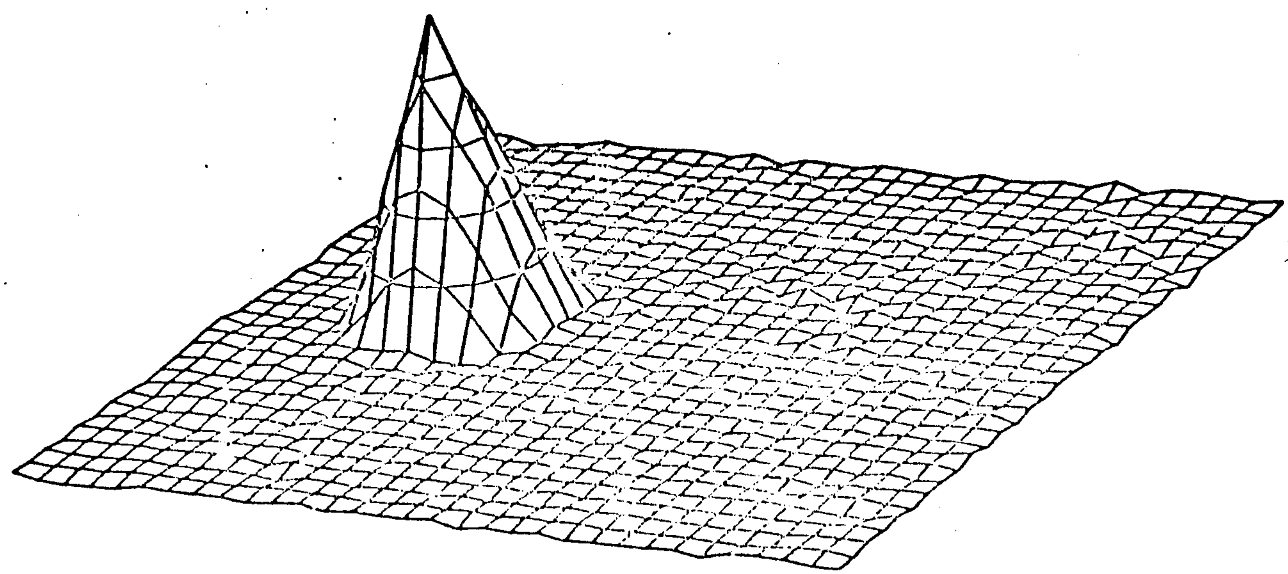


Fig. 11.19 Same as Fig. 11.17 except using a Fourier-spectral method with $K = P = 16$ (32 x 32 modes).

$$u(\pm 1,t) = 0 \tag{11.16}$$

$$u(x,0) = f(x),\ \frac{\partial u}{\partial t}(x,0) = g(x) \tag{11.17}$$

is well posed. Legendre polynomial solution of (11.15-17) is semi-bounded and, hence stable (see Sec. 7). However, we have not yet been able to prove that Chebyshev solution of this problem is ever algebraically stable. The techniques of Sec. 8 prove that if the boundary conditions (11.16) are replaced by the characteristic boundary conditions

$$\frac{\partial u(-1,t)}{\partial t} - \frac{\partial u(-1,t)}{\partial x} = 0, \quad \frac{\partial u(1,t)}{\partial t} + \frac{\partial u(1,t)}{\partial x} = 0,$$

the scheme is algebraically stable. However, we have not yet been able to prove this result for the boundary conditions (11.16). It is reassuring to note that we have solved the Cheybshev-spectral approximations to (11.15-17) and find no evidence of lack of convergence. Indeed, the Chebyshev methods work just as well as they do for (11.1-2). Thus, it is not the spectral methods that run into difficulty on higher-order equations, but just our methods of analysis.

12. Advective-Diffusion Equation

In this section, we consider spectral methods for the advective-diffusion ('linearized Burgers') equation

$$\frac{\partial u(x,t)}{\partial t} + U\frac{\partial u(x,t)}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2} + f(x,t) \qquad (-1 \le x \le 1) \tag{12.1}$$

$$u(-1,t) = 0\,, \qquad u(1,t) = 0 \tag{12.2}$$

$$u(x,0) = g(x) \tag{12.3}$$

Eq. (12.1) is parabolic so boundary conditions should be applied at both $x = -1$ and $x = +1$. When ν is small, the boundary condition applied at $x = +1$ (assuming $U > 0$) has an interesting effect on the stability of the spectral methods.

To begin, we remark that the analyses of Sects. 7-8 can be extended to show that, as $N \to \infty$, N-term Legendre and Chebyshev approximations to (12.1-3) are stable and convergent. For example, Chebyshev-Galerkin approximation is stable because (12.1-2) and (7.3) imply that

$$\begin{aligned}\frac{d}{dt}\int_{-1}^{1}\frac{u^2}{\sqrt{1-x^2}}\,dx &\le |U|\int_{-1}^{1}\frac{u^2}{(1-x^2)^{3/2}}\,dx - \nu\int_{-1}^{1}\frac{u^2}{(1-x^2)^{5/2}}\,dx\\ &\le |U|\int_{-\sqrt{1-\nu/U}}^{\sqrt{1-\nu/U}}\frac{u^2}{(1-x^2)^{3/2}}\,dx\\ &\le \frac{U^2}{\nu}\int_{-1}^{1}\frac{u^2}{\sqrt{1-x^2}}\,dx\end{aligned} \tag{12.4}$$

so the approximation is semi-bounded.

However, for finite N, there may be difficulty integrating the resulting spectral equations. With Legendre polynomials, Galerkin approximation L_N to (12.1-3) satisfies $L_N + L_N^* \leq 0$ so there is no difficulty with time integrations (although the solution may not be accurate unless N is large enough.

On the other hand, Chebyshev-spectral solution of (12.1-3) encounters the following curious behavior when ν is small. If ν/U is small and N is not too large, the Chebyshev-spectral approximations L_N to (12.1-3) have eigenvalues with positive real parts. In Table 12.1, we list values of N_{crit} for various values of ν/U; for $N < N_{crit}$, L_N for Chebyshev-tau approximation to (12.1-3) has eigenvalues with positive real parts. Since these eigenvalues may have moderately large real parts [they can be as large as $U^2/2\nu$ by (12.4)], there may be rapid growth of errors and numerical solution of the Chebyshev-spectral equations may appear unstable and divergent. For $N \geq N_{crit}$, there are no eigenvalues of L_N with positive real parts so the spectral equations are stable.

The origin of this temporal instability is the outflow boundary layer at $x = \pm 1$; when $U > 0$, the solution to (12.1-3) develops a region of rapid change of width roughly ν/U near $x = +1$ as t increases. Since roughly $3(U/\nu)^{1/2}$ Chebyshev polynomials are required to resolve a boundary layer of width ν/U [see (3.50)], we expect that $N_{crit} \doteq 3(U/\nu)^{1/2}$ so $\nu N_{crit}^2/U \doteq 9$. In fact, as shown in Table 12.1, the criterion is actually $\nu N_{crit}^2/U \doteq 4$. [Since the Chebyshev norm of $\exp(-Ut\partial/\partial x)$ is roughly $N^{1/4}$ (see Sec. 8), we expect that the proper scaling of N_{crit} is better represented as $\nu N_{crit}^{7/4}/U \doteq 1.3$. As shown in Table 12.1, this modified scaling is more nearly satisfied for the range of ν considered.]

TABLE 12.1

ν/U	N_{crit}	$\nu N_{crit}^2/U$	$\nu N_{crit}^{7/4}/U$
1.0×10^{-2}	15	2.25	1.14
2.5×10^{-3}	35	3.06	1.26
1.0×10^{-3}	61	3.72	1.33
6.0×10^{-4}	81	3.94	1.31
4.0×10^{-4}	101	4.08	1.29

Table 12.1 Critical values N_{crit} of the number of Chebyshev polynomials necessary that the tau approximation to the operator $-U\partial u/\partial x + \nu\partial^2 u/\partial x^2$ with $u(\pm 1) = 0$ have no eigenvalue with positive real parts. Also listed are the inverse 'grid Reynolds number' $\nu N_{crit}^2/U$ and the parameter $\nu N_{crit}^{7/4}/U$.

If Chebyshev-spectral approximations to (12.1-3) are solved using fractional time-step methods, the temporal instability for $N < N_{crit}$ appears in a unique way. Define the operator A_N as an N-mode Chebyshev approximation to the operator $-U\partial u/\partial x$ with the boundary condition $u(-1) = 0$ and the operator B_N as an N-mode Chebyshev approximation to the operator $\nu\partial^2 u/\partial x^2$ with $u(\pm 1) = 0$. Then the evolution operator of (12.1-2) is $\exp[(A_N+B_N)t]$ so a fractional step method involves the splitting $\partial u_N/\partial t = \partial_1 u_N/\partial t + \partial_2 u_N/\partial t$ where

$$\partial_1 u_N/\partial t = A_N u_N , \qquad \partial_2 u_N/\partial t = B_N u_N .$$

For any values of ν and $U > 0$, the fractional step $\partial_1 u_N/\partial t$ is algebraically stable since $\|\exp A_N t\| = O(N^{1/4})$ (see Sec. 8), while the fractional step $\partial_2 u_N/\partial t$ is stable since $\|\exp B_N t\| \leq 1$ (see Sec. 7). Nevertheless, $\|\exp[(A_N+B_N)t]\|$ can grow rapidly with t. The reason is that A_N and B_N do not commute so it is not true that $\|\exp[(A_N+B_N)t]\| \leq \|\exp A_N t\| \; \|\exp B_N t\|$.
The Lie formula (5.8) does ensure that

$$\|\exp[(A_N+B_N)t]\| \leq \lim_{n\to\infty} \|\exp(A_N t/n)\|^n \; \|\exp(B_N t/n\|^n .$$

However, as $N \to \infty$ with $n/N^2 \to \infty$,

$$\|\exp(A_N t/n)\| - 1 \sim cN^2 t/n$$

with $c > 0$ (see Table 8.1) so

$$\|\exp(A_N t/n)\|^n \sim \exp(cN^2 t) \gg 1 \quad (N \to \infty, \; n/N^2 \to \infty) .$$

Therefore the Lie formula gives only the very weak upper bound

$$\|\exp[(A_N + B_N)t]\| \leq \exp(cN^2 t) .$$

In summary, Chebyshev-spectral approximations to (12.1-3) give fractional step methods such that each fractional step is algebraically stable while the total step is unstable unless $N > N_{crit}$.

If the boundary conditions (12.2) are replaced by

$$u(-1,t) = 0, \qquad \frac{\partial u}{\partial x}(+1,t) = 0 \tag{12.4}$$

when $U > 0$, the criterion for temporal stability is relaxed significantly. As shown in Table 12.2, the value of $\nu N^2_{crit}/U$ is decreased to roughly 1.6. However, the growing modes that appear when $N < N_{crit}$ are much tamer than those that appear when the boundary condition $u(+1,t) = 0$ is applied, so accurate time integrations are still practicable when $\nu N^2/U \gtrsim 0.01$ (see Haidvogel 1977).

TABLE 12.2

ν/U	N_{crit}	$\nu N_{crit}^{7/4}/U$
2.5×10^{-3}	21	0.52
1.0×10^{-3}	37	0.56
6.0×10^{-4}	49	0.54
4.0×10^{-4}	61	0.53
2.0×10^{-4}	89	0.52

Table 12.2. Critical values N_{crit} of the number of Chebyshev polynomials necessary that the tau approximation to the operator $-U\partial u/\partial x + \nu\partial^2 u/\partial x^2$ with $u(-1) = 0$, $\partial u(+1)/\partial x = 0$ and $U > 0$ have no eigenvalues with positive real parts. The parameter $\nu N_{crit}^{7/4}/U$ is also listed.

13. Models of Incompressible Fluid Dynamics

The Stokes equations for low Reynolds number, two-dimensional incompressible flow are

$$\begin{aligned} \frac{\partial \vec{v}}{\partial t} &= -\vec{\nabla} p + \nu \nabla^2 \vec{v}, \\ \vec{\nabla}\cdot\vec{v} &= 0 \ , \end{aligned} \tag{13.1}$$

where $\vec{v}$ is the velocity field, p is the pressure, and ν is the kinematic viscosity. With the boundary conditions that $\vec{v} = 0$ on rigid stationary boundaries, the problem (13.1) is well posed for any $\nu > 0$. An equivalent formulation is given by the vorticity-streamfunction equations

$$\begin{aligned} \frac{\partial \zeta}{\partial t} &= \nu \nabla^2 \zeta \ , \\ \zeta &= \nabla^2 \psi \ , \end{aligned} \tag{13.2}$$

obtained by taking the curl of the Stokes equations (13.1). Here ψ is the streamfunction defined by $\vec{v} = (-\partial\psi/\partial y, \partial\psi/\partial x)$ and ζ is the vorticity.

A one-dimensional model of (13.2) is

$$\frac{\partial \zeta}{\partial t} = \nu \frac{\partial^2 \zeta}{\partial x^2} \quad (-1 \leq x \leq 1,\ t > 0), \tag{13.3}$$

$$\zeta = \frac{\partial^2 \psi}{\partial x^2} \quad (-1 \leq x \leq 1). \tag{13.4}$$

On stationary rigid walls, the boundary conditions for (13.3-4) are

$$\psi(x,t) = \psi_x(x,t) = 0 \quad (x = \pm 1). \tag{13.5}$$

There is one subtlety in the application of spectral methods to (13.3-5) that does not appear directly when the primitive equations (13.1) are used. It is necessary to use some care to avoid unconditional numerical instability with the Chebyshev-tau method.

The most obvious way to use the tau method to solve (13.3-5) is to substitute (13.4) into (13.3) and solve

$$\psi_{xxt} = \nu\psi_{xxxx} \quad (-1 \le x \le 1,\ t > 0) \tag{13.6}$$

by expanding $\psi(x,t)$ in the Chebyshev series

$$\psi(x,t) = \sum_{n=0}^{N} a_n(t)T_n(x) . \tag{13.7}$$

Denoting by $a_n^{(q)}$ the Chebyshev expansion coefficients of $\partial^q\psi/\partial x^q$ (see A.23), the tau equations for (13.5-6) are

$$\frac{da_n^{(2)}}{dt} = \nu\ a_n^{(4)} \quad (0 \le n \le N-4,\ t > 0), \tag{13.8}$$

$$\sum_{n=0}^{N} (\pm 1)^n a_n = \sum_{n=0}^{N} (\pm 1)^n n^2 a_n = 0. \tag{13.9}$$

Unfortunately, this method for solution of (13.3-5) is unconditionally unstable as $N \to \infty$. In Table 13.1, we list the largest positive eigenvalue λ_{max} of (13.8-9); there is a solution of (13.8-9) that grows like $a_n(t) = a_n(0)\exp(\lambda_{max}t)$. Since λ_{max} grows like N^4 as $N \to \infty$, errors also grow rapidly as $N \to \infty$ for fixed t. This method is unusable for time-dependent calculations.

In Table 13.1, we also list the values of λ_n for n = 1,5, where the eigenvalues of (13.8-9) are ordered according to $|\lambda_1| \le |\lambda_2| \le \ldots$. The exact eigenvalues of (13.3-5) are found by seeking solutions of these equations of the form $\psi(x,t) = \psi(x)\exp(\lambda t)$, $\zeta(x,t) = \zeta(x)\exp(\lambda t)$. It may be easily verified that the exact eigenvalues of (13.3-5) are given by $\lambda = -\mu^2$ with $\mu = n\pi$ or μ any nonzero solution of the transcendental equation $\tan\mu = \mu$. The exact values of λ_1 and λ_5 are also listed

Table 13.1

N	λ_1	λ_5	λ_{max}
10	-9.8696598	-189.63800	4,272.
15	-9.8696044	- 89.54550	29,439.
20	-9.8696044	- 88.86244	111,226.
25	-9.8696044	- 88.86244	294,697.
30	-9.8696044	- 88.86244	652,722.
35	-9.8696044	- 88.86244	1,255,298.
40	-9.8696044	- 88.86244	2,215,880.
Exact	-9.8696044	- 88.86244	

Table 13.1. Eigenvalues of the tau approximation (13.8-9) to (13.6-7). The N-4 eigenvalues are ordered so that $|\lambda_1| \leq |\lambda_2| \leq \ldots \leq |\lambda_{N-4}|$. All the eigenvalues are real. The largest positive eigenvalue $\lambda_{max} = \lambda_{N-4}$.

in Table 13.1. Evidently, even though (13.8-9) is unstable as $N \to \infty$, it does a good job of reproducing the low-n modes; approximately $\sqrt{|\lambda_n|}$ Chebyshev polynomials are required to resolve the mode with eigenvalue λ_n. Thus, this version of the tau method may be useful for eigenvalue calculations even though it is unconditionally unstable for the initial-value problem (13.3-5) (as evidenced by the spurious unstable modes with eigenvalues as large as λ_{max}).

The tau method behaves similarly when applied to more difficult problems, like the Orr-Sommerfeld equation for linear stability analysis of incompressible plane-parallel shear flows. Low modes are given accurately by the analog of (13.8-9) (see Orszag 1971d), but there appear spurious unstable modes with large growth rates. Similar spurious unstable modes appear in finite-difference solution of the Orr-Sommerfeld equation (see Gary & Helgason 1970).

There is a simple method to avoid the spurious unstable modes encountered by (13.8-9). The technique to be described below also eliminates the spurious unstable modes encountered in solution of the Orr-Sommerfeld equation. The idea is simply not to combine (13.3-4) into (13.6). Rather, we expand $\zeta(x,t)$ as in

$$\zeta(x,t) = \sum_{n=0}^{N} b_n(t) T_n(x) \tag{13.10}$$

and solve

$$\frac{db_n}{dt} = \nu\, b_n^{(2)} \qquad (0 \le n \le N-2), \tag{13.11}$$

$$b_n = a_n^{(2)} \qquad (0 \le n \le N-2), \tag{13.12}$$

in addition to (13.9). Here we have dropped two equations from the Chebyshev modal equations that result from (13.3-4). The logic of this modification of the tau method is as follows. Application of (13.8) for $0 \leq n \leq N-4$ is equivalent to application of (13.12) for $0 \leq n \leq N$ together with (13.11) for $0 \leq n \leq N-4$. On physical grounds, we may expect that this procedure will lead to instability because the boundary conditions $\psi = 0$ at $x = \pm 1$ should be imposed on (13.4) not (13.3), while the boundary conditions $\psi_x = 0$ at $x = \pm 1$ should be imposed on (13.3) only when $\nu > 0$. On the other hand, when the system is truncated as in (13.11-12), each of the dynamical equations can play their proper role in adjusting the boundary conditions: the boundary conditions $\psi = 0$ are imposed on (13.12) while the boundary conditions $\psi_x = 0$ are imposed on (13.11).

We shall now prove that (13.11-12) is stable for the special case in which N is even with $a_{2n+1} = b_{2n+1} = 0$ for all n, $t \geq 0$. In this case, $\psi(x,t)$ and $\zeta(x,t)$ are even functions of x. To begin , we observe that (13.11) is equivalent to

$$\frac{\partial \zeta}{\partial t} = \nu \frac{\partial^2 \zeta}{\partial x^2} + b_N' T_N(x) \qquad (-1 \leq x \leq 1,\ t > 0) \ ,$$

while (13.12) is equivalent to

$$\zeta(x,t) = \frac{\partial^2 \psi}{\partial x^2} + b_N T_N(x) \ .$$

Therefore,

$$\frac{\partial^3}{\partial t \partial x^2} \psi = \nu \frac{\partial^4 \psi}{\partial x^4} + b_N T_N'' \ .$$

Since ψ is an even function of x, it follows by integration with respect to x that

$$\frac{\partial^2 \psi}{\partial x \partial t} = \nu \frac{\partial^3 \psi}{\partial x^3} + b_N T_N' \ . \qquad (13.13)$$

Also, since $\psi(x,t)$ is a polynomial of degree N that satisfies $\psi_x(\pm 1,t) = 0$, integration by parts gives

$$\int_{-1}^{1} \psi_x T_N'(1-x^2)^{-\frac{1}{2}}dx = -\int_{-1}^{1} [\psi_{xx} + x\psi_x/(1-x^2)]T_N(1-x^2)^{-\frac{1}{2}}dx = 0$$

since ψ_{xx} and $x\psi_x/(1-x^2)$ are polynomials of degree N-2 so they must be orthogonal to $T_N(x)$. Therefore, taking the Chebyshev inner product of (13.13) and $\psi_x(x,t)$, we obtain

$$\frac{\partial}{\partial t} \int_{-1}^{1} \psi_x^2(1-x^2)^{-\frac{1}{2}}dx = 2\nu \int_{-1}^{1} \psi_x\psi_{xxx}(1-x^2)^{-\frac{1}{2}}dx \leq 0, \quad (13.14)$$

where the last inequality is established using the inequality derived in Example 7.1(v):

$$\int_{-1}^{1} uu_{xx}(1-x^2)^{-\frac{1}{2}}dx \leq 0,$$

if $u(x)$ is a polynomial of degree N satisfying $u(\pm 1) = 0$. The energy bound (13.14) proves stability of the tau approximation (13.11-12).

Finally, let us discuss methods for the solution of the primitive equations (13.1) using Chebyshev tau approximations. A one-dimensional model that embodies the essential features of (13.1) is obtained by solving (13.1) within the slab $-1 \leq x \leq 1$, $-\infty \leq y \leq \infty$, with an assumed solution of the form

$$\vec{v} = (u(x,t)e^{iky}, v(x,t)e^{iky}), \; p = p(x,t)e^{iky}$$

for some real wavenumber k. Let the Chebyshev expansion coefficients of $u(x,t)$, $v(x,t)$, $p(x,t)$ be denoted as $u_n(t)$, $v_n(t)$, $p_n(t)$ $(0 \leq n \leq N)$, respectively. Then an unconditionally stable, implicit fractional step method for the solution of (13.1) with a forcing term $(f(x,t)e^{iky}, g(x,t)e^{iky})$ added to the right side is

$$\bar{u}_n = u_n(t) + \Delta t[-p_n^{(1)} + f_n(t)] \quad (0 \le n \le N-2), \qquad (13.15)$$

$$\bar{v}_n = v_n(t) + \Delta t[-ikp_n + g_n(t)] \quad (0 \le n \le N), \qquad (13.16)$$

$$\bar{u}_n^{(1)} + ik\bar{v}_n = 0 \quad (0 \le n \le N), \qquad (13.17)$$

$$\sum_{n=0}^{N} \bar{u}_n = \sum_{n=0}^{N} (-1)^n \bar{u}_n = 0 \quad (0 \le n \le N), \qquad (13.18)$$

$$u_n(t+\Delta t) - \nu\Delta t u_n^{(2)}(t+\Delta t) = \bar{u}_n \quad (0 \le n \le N-2), \qquad (13.19)$$

$$v_n(t+\Delta t) - \nu\Delta t v_n^{(2)}(t+\Delta t) = \bar{v}_n \quad (0 \le n \le N-2), \qquad (13.20)$$

$$\sum_{n=0}^{N} (\pm 1)^n u_n(t+\Delta t) = \sum_{n=0}^{N} (\pm 1)^n v_n(t+\Delta t) = 0. \qquad (13.21)$$

Here we use the notation that, for example, $u_n^{(2)}$ represents the Chebyshev coefficients of $u_{xx}(x,t)$. The scheme (13.15-21) is based on backwards Euler time differencing; it is straightforward to generalize (13.15-21) to other more accurate time differencing methods.

The fractional step (13.15-18) involves computation of the pressure field by imposition of the incompressibility condition (13.17). Only the boundary conditions $u(\pm 1,t) = 0$ are applied because this part of the time step is effectively inviscid so only the normal flow can be specified at the boundary. Thus, we drop (13.15) for $n = N-1,N$ in favor of the two boundary conditions (13.18). The fractional step (13.19-21) involves the viscous term in (13.1) so boundary conditions are applied on both the normal velocity component u and the tangential velocity component v. Accordingly, the tau method involves dropping (13.19-20) for $n = N-1,N$ in favor of these boundary conditions.

The system (13.15-21) is solved as follows. Multiplying (13.15) by ik and subtracting the result from the Chebyshev x-derivative of (13.16) gives

$$\bar{v}_n^{(1)} - ik\bar{u}_n = v_n^{(1)}(t) - iku_n(t) + \Delta t[g_n^{(1)}(t) - ikf_n(t)]$$
$$(0 \le n \le N-2).$$

Substituting $\bar{v}_n = i\bar{u}_n^{(1)}/k$ from (13.17) gives

$$\bar{u}_n^{(2)} - k^2\bar{u}_n = u_n^{(2)}(t) - k^2u_n(t) + \Delta t[-ikg_n^{(1)}(t) - k^2f_n(t)]$$
$$(0 \le n \le N-2). \quad (13.22)$$

Eq. (13.22) with the boundary conditions (13.18) is of the same form as (13.19-20) with boundary conditions (13.21). These equations are best solved by the algorithm discussed at the end of Sec. 10.

The stability analysis of (13.15-21) is as follows. The evolution of a perturbation is governed by (13.15-21) with $f_n = g_n = 0$ for all n. Therefore, the solution of (13.22) is $\bar{u}_n = u_n(t)$ for all n. Also, $\bar{v}_n = v_n(t)$ for all n. Finally, the implicit scheme (13.19-21) is an unconditionally stable scheme for solution of the heat equation. This proves that (13.15-21) is unconditionally stable.

Implicit, unconditionally stable methods for solution of (13.1) that do not use fractional steps [as in (13.15-21)] can also be implemented. In a fully implicit scheme, we would solve

$$u_n(t+\Delta t) = u_n(t) + \Delta t[-p_n^{(1)} + \nu u_n^{(2)}(t+\Delta t) + f_n(t+\Delta t)]$$
$$(0 \le n \le N-2), \quad (13.23)$$

$$v_n(t+\Delta t) = v_n(t)+\Delta t[-ikp_n+\nu v_n^{(2)}(t+\Delta t)+g_n(t+\Delta t)]$$

$$(0\leq n\leq N-2), \qquad (13.24)$$

$$u_n^{(1)}(t+\Delta t)+ikv_n(t+\Delta t) = 0 \quad (0\leq n\leq N), \qquad (13.25)$$

$$\sum_{n=0}^{N} (\pm 1)^n u_n(t+\Delta t) = \sum_{n=0}^{N} (\pm 1)^n v_n(t+\Delta t) = 0. \qquad (13.26)$$

Substituting (13.25) into (13.24) and eliminating p_n between (13.23 - 24), we obtain

$$u_n^{(2)}(t+\Delta t) - k^2 u_n(t+\Delta t) = u_n^{(2)}(t)-k^2 u_n(t)$$

$$+\Delta t[\nu u_n^{(4)}(t+\Delta t)-\nu k^2 u_n^{(2)}(t+\Delta t)-ikg_n^{(1)}-k^2 f_n]$$

$$+b_1+b_2(-1)^n \qquad (0\leq n\leq N-2), \qquad (13.27)$$

$$\sum (\pm 1)^n u_n(t+\Delta t) = \sum (\pm 1)^n n^2 u_n(t+\Delta t) = 0. \qquad (13.28)$$

Here b_1 and b_2 are parameters determined by the condition that (13.27-28) have a solution.

There are two features of the full implicit scheme (13.23-28) that should be mentioned in comparison with the fractional step scheme (13.15-22). First, the solution of (13.27-28) involves solution of an essentially pentadiagonal matrix equation in contrast to (13.22) which is essentially tridiagonal. This may be a serious difficulty because the pentadiagonal system (13.27) is not as well conditioned as the tridiagonal system (13.22).

Second, the full implicit scheme avoids an ambiguity of the Navier-Stokes equations pointed out by Orszag & Israeli (1974, p. 299). When the boundary conditions $\vec{v} = 0$ are applied to (13.1) (with a force term included), we obtain

$$\vec{\nabla} p = \nu \nabla^2 \vec{v} + \vec{f} \tag{13.29}$$

on the boundaries. Therefore, we can obtain boundary conditions on both $\partial p/\partial n$ and p, where n is normal to the boundary. However, the divergence of (13.1) gives

$$\nabla^2 p = \vec{\nabla} \cdot \vec{f}, \tag{13.30}$$

so p is the solution of the Poisson equation (13.30) satisfying both the Dirichlet and Neumann boundary conditions (13.29). It seems at first that p is overspecified. In fact, p is not overspecified; the above argument reflects only the adjustment that $\vec{v}$ must undergo near the boundary in order to make both sets of boundary conditions on p mutually consistent.

This adjustment process is directly accounted for in the system (13.23-28) but is only indirectly accounted for in the fractional step method (13.15-22). Only the boundary condition (13.18) is applied while determining the pressure in (13.15-17). Nevertheless, it seems that the fractional step method adjusts itself from time step to time step so no serious errors are produced by neglecting the tangential components of (13.29).

The methods discussed in this section extend to give stable methods for solution of the nonlinear Navier-Stokes equations. For example, if the forcing term (f,g) in (13.15-16) is chosen to be the nonlinear terms of the Navier-Stokes equations, our analysis shows that stability of (13.15-21) is determined by stability restrictions on the nonlinear terms alone.

14. Miscellaneous Applications of Spectral Methods

In this Section, we survey some special topics regarding spectral methods. Some of these topics are still under active investigation, so the results reported here are very incomplete.

Complicated Geometries

There are two ways that spectral methods can be used to solve problems in complicated geometries without introducing basis functions that are special to the geometry and, therefore, unwieldy and inefficient to use. The two methods are mapping and patching.

Mapping involves transforming the complicated domain into a simpler one by means of a coordinate transformation. Spectral methods are then applied in the simple geometry using the techniques discussed in earlier sections. For example, if we wish to solve the heat equation

$$\frac{\partial}{\partial t} u(x,y,t) = \nabla^2 u(x,y,t) \tag{14.1}$$

in the two-dimensional domain

$$-1 \leq x \leq 1, \quad -f(x) \leq y \leq f(x)$$

for some given function $f(x)$ with the boundary conditions that $u = 0$ on the boundary of the domain, we would proceed as follows. First, we make the coordinate transformation

$$z = y/f(x) \qquad (-1 \leq z \leq 1) \tag{14.2}$$

and rewrite (14.1) as

$$\frac{\partial}{\partial t} u(x,z,t) = \left(\frac{\partial}{\partial x} - \frac{f'}{f} z \frac{\partial}{\partial z}\right)^2 u(x,z,t) + f^{-2} \frac{\partial^2}{\partial z^2} u(x,z,t)$$

$$(-1 \leq x \leq 1, \ -1 \leq z \leq 1) \ . \tag{14.3}$$

Then, we expand $u(x,z,t)$ in a double Chebyshev series and integrate (14.3). For this purpose, a hybrid numerical scheme is recommended. Unconditionally stable time differencing can be obtained by a semi-implicit method (see Sec. 10). Here a simple diffusion operator is added and subtracted from (14.3). The simple diffusion operator is then evaluated implicitly using tau method (because the tau method is simplest when no complicated nonlinearities or nonconstant coefficient terms are involved); the remaining nonconstant coefficient term in (14.3) is then evaluated explicitly using fast Fourier transforms and the collocation method. The result is an efficient and accurate method for solution of (14.1).

Techniques like those just described have been applied to a variety of problems with much success. If a convenient coordinate transformation is available, the mapping technique combined with appropriate spectral methods may be expected to be very useful.

The idea of patching is that if the geometry is the union of several simpler geometries (like an L-shaped region) then spectral approximations can be formulated in each of the simpler domains. These solutions are then patched across the boundaries by requiring that the solution (and an appropriate number of derivatives) be smooth. When this technique is applied together with the mapping technique discussed above, it is possible to devise spectral shock-fitting methods for the solution of compressible flow problems. Patching methods require much further investigation to judge their usefulness in practical problems.

Poisson's Equation in Two and Higher Dimensions

The Chebyshev tau equations for Poisson's equation $\nabla^2 u = f$ in the square $-1 \le x \le 1$, $-1 \le y \le 1$ are

$$u_{nm}^{(2,0)} + u_{nm}^{(0,2)} = f_{nm} \quad (0 \le n \le N-2, 0 \le m \le M-2), \tag{14.4}$$

while the Dirichlet boundary conditions u = 0 are

$$\sum_{n=0}^{N} (\pm 1)^n u_{nm} = 0 \qquad (0 \le m \le M) \quad , \tag{14.5}$$

$$\sum_{m=0}^{N} (\pm 1)^m u_{nm} = 0 \qquad (0 \le n \le N) \quad . \tag{14.6}$$

Here we expand $u(x,y)$ and $f(x,y)$ in the double Chebyshev series

$$\begin{Bmatrix} u(x,y) \\ f(x,y) \end{Bmatrix} = \sum_{n=0}^{N} \sum_{m=0}^{M} \begin{Bmatrix} u_{nm} \\ f_{nm} \end{Bmatrix} T_n(x) T_m(y) \tag{14.7}$$

and we denote the Chebyshev expansion coefficients of $\partial^{p+q} u/\partial x^p \partial y^q$ by $u_{nm}^{(p,q)}$. The $2N+2M+4$ boundary conditions are not all linearly independent; there exist four linear relations among them, namely

$$\sum_{n=0}^{N} \sum_{m=0}^{M} (\pm 1)^n (\pm 1)^m u_{nm} = 0. \tag{14.8}$$

Thus, (14.4-6) gives $(N+1)(M+1)$ equations for the $(N+1)(M+1)$ unknowns u_{nm} $(0 \le n \le N,\ 0 \le m \le M)$.

Using (10.10) [or (A.23)], the system (14.4-6) can be reduced to a block tridiagonal matrix equation modified by extra full rows corresponding to the boundary conditions (14.5-6). These equations can be solved by standard block tridiagonal algorithms in order $N^3 M$ or order NM^3 operations. If Poisson's equation must be solved several times with the same values of N and M but different functions $f(x,y)$, it is more efficient to apply alternative methods.

A method to solve (14.4-6) in order N^2M operations (with a preprocessing stage that requires order N^3 operations) is as follows. First, we find the N-2 eigenvalues λ_p and eigenvectors e_{np} $(p = 0,\ldots,N-2)$ of the equations

$$e_{np}^{(2)} = \lambda_p e_{np} \quad (0 \le n \le N-2)$$

$$\sum_{n=0}^{N} (\pm 1)^n e_{np} = 0.$$

The eigenvalues λ_p are all negative as proved in Example 7.3(ii). Then we form the $(N+1)\times(N+1)$ matrix E whose elements are

$$E_{np} = e_{np} \quad (0 \le n \le N,\ 0 \le p \le N-2)$$
$$E_{n,N-1} = \delta_{n,0} \quad (0 \le n \le N)$$
$$E_{n,N} = \delta_{n,1} \quad (0 \le n \le N)$$

and compute the inverse matrix $D = E^{-1}$. Since the boundary conditions (14.5) are satisfied by u_{nm}, it follows that

$$u_{nm} = \sum_{p=0}^{N-2} e_{np} v_{pm} \tag{14.9}$$

for suitable v_{pm} for all n,m. Therefore, setting

$$g_{pm} = \sum_{n=0}^{N} (D)_{pn} f_{nm} \quad (0\le p\le N-2,\ 0\le m\le M-2), \tag{14.10}$$

it follows that (14.4-6) become

$$\lambda_p v_{pm} + v_{pm}^{(0,2)} = g_{pm} \quad (0\le p\le N-2,\ 0\le m\le M-2) \tag{14.11}$$

$$\sum_{m=0}^{M} (\pm 1)^m v_{pm} = 0 \quad (0\le p\le N-2). \tag{14.12}$$

Eqs. (14.11-12) may be solved efficiently (in order NM operations) for v_{pm} using the algorithm discussed at the end of Sec. 10. Once v_{pm} is found, u_{nm} may be reconstructed from (14.9). The total operation count is order N^2M [from the two matrix multiplies

(14.9-10)].

The solution of Poisson's equation by the Chebyshev series method outlined above is very competitive with finite-difference solution using fast Poisson solvers. Zang & Haidvogel (1977) present a number of comparisons of the Chebyshev methods and fast Poisson solvers.

There are two further complications that may arise in elliptic boundary-value problems. First, the elliptic equation may have nonconstant coefficients or may even be nonlinear. Here we recommend that spectral equations be solved using relaxation methods of the kind advocated by Concus & Golub (1973), in which the heart of the algorithm is the fast, efficient solution of Poisson-like equations. Second, the geometry may be more complicated than a box. In this case, we recommend the implementation of capacitance matrix techniques (or equivalent Green's function techniques) in which the problem to be solved is imbedded in a simpler geometry, like a box (see Buzbee et al 1971). Again, the heart of the algorithm is the fast solution of Poisson's equation using (14.9-12).

Coordinate Singularities

When spectral methods are applied to problems in cylindrical or spherical geometries, their formualtion may require special care at the coordinate singularities. These 'pole problems' have been extensively investigated (Orszag 1974, Tang 1977). As a simple example of these effects, let us consider the computation of the eigenvalues of Bessel's equation using the Chebyshev tau method (Metcalfe 1974). The problem is to find the eigenvalues and eigenfunctions y(x) of

$$y'' + \frac{1}{x} y' - \frac{n^2}{x^2} y = - \lambda y \qquad (14.13)$$

subject to the conditions that y(1) = 0 and that y(x) be finite for $0 \le x \le 1$. The exact eigenvalues are related to the zeros of the Bessel function J_n: $\lambda_p = j_{np}^2$ where $J_n(j_{np})=0$, p=1,2,... .

When n is even, the eigenfunctions of (14.13) are even functions of x; when n is odd, the eigenfunctions are odd. This fact suggests that we represent the solution to (14.13) in terms of series of even Chebyshev polynomials when n is even and odd polynomials when n is odd. Thus, for n odd we write

$$y(x) = \sum_{m=1}^{M} y_m T_{2m-1}(x) \quad . \qquad (14.14)$$

In Table 14.1, we list numerical values for the smallest eigenvalue of (14.13) with n = 7 using the series (14.14), the boundary condition y(1) = 0, and the Chebyshev tau method. The convergence of this method, while very impressive as M increases, is slowed by the coordinate singularity of (14.13) at x = 0. In general, series of the form (14.14) behave like x as $x \to 0$. In this case the terms y'/x and y/x^2 are singular at x = 0. The true eigenfunctions $J_7(j_{7p}x)$ behave like x^7 as $x \to 0$, as may easily be shown using Frobenius' method, so none of the terms of (14.13) are in fact singular for the exact eigenfunctions.

It is possible to improve the convergence of (14.14) by imposing additional 'pole conditions', like y'(0) = 0. When y'(0) = 0 in the series (14.14), the terms of (14.13) are individually nonsingular. In Table 14.1, we also list numerical values of the smallest eigenvalue of (14.13) with n = 7 and the two boundary conditions y(1) = 0, y'(0) = 0 applied. There

Table 14.1

M	λ_1 with $y(1)=0$	λ_1 with $y(1)=y'(0)=0$
10		124.001290649
14	169.111983340	122.895944051
18	126.557832251	122.907620295
22	122.991799598	122.907600279
26	122.908250800	122.907600204
Exact	122.907600204	122.907600204

Table 14.1. Smallest eigenvalue of (14.13) with $n = 7$ obtained using (14.14) and the Chebyshev tau method. M is the number of Chebyshev polynomials. The extra boundary condition $y'(0) = 0$ is a pole constraint at the singular point $x = 0$ of (14.13).

is clearly a dramatic improvement in the rate of convergence. It is also possible to make the problem less sensitive to pole properties near the origin by first multiplying (14.13) by x^2 to eliminate explicitly singular terms and then applying the tau method. The results of the latter trick are essentially the same as applying the pole condition $y'(0) = 0$ directly to (14.13).

If pole conditions are not properly applied, it is possible to degrade significantly the accuracy of spectral computations. It is even possible to induce strong instabilities that are absent when proper pole conditions are applied. These matters are discussed in detail by Orszag (1974) and Tang (1977).

Energy Conservation

It was shown in Sec. 2 (see footnote on p.29) that if $(u,A(u)) = 0$ so the solution to the nonlinear equation $\partial u/\partial t = A(u)$ conserves energy $[\partial(u,u)/\partial t = 0]$, then the solution to any spectral approximation obtained by a self-adjoint projection operator also conserves energy. Some examples of this result are energy conservation by Galerkin approximations to the inviscid Navier-Stokes equations [(9.55) with $\nu = 0$] using Fourier series with periodic or free-slip boundary conditions and Legendre polynomial series with rigid no-slip boundary conditions.

If the inviscid Navier-Stokes equations are rewritten in so-called rotational form as

$$\frac{\partial \vec{v}}{\partial t} = \vec{v} \times \vec{\omega} - \nabla \Pi + \nu \nabla^2 \vec{v} , \tag{14.15}$$

where $\vec{\omega} = \vec{\nabla} \times \vec{v}$ is the vorticity and $\Pi = p + \frac{1}{2} v^2$ is the pressure head, then energy conservation holds when $\nu = 0$ for certain collocation approximations. If the collocation points are $\vec{x}_i$ and $\vec{v}_i = \vec{v}(\vec{x}_i)$, then (14.15) gives

$$\frac{\partial}{\partial t} \sum_i v_i^2 = - \sum_i \vec{v}_i \cdot \vec{\nabla} \Pi_i \tag{14.16}$$

when $\nu = 0$. If the collocation projection operator is such that integration by parts is valid in the sense that

$$\sum_i \vec{v}_i \cdot \vec{\nabla} \Pi_i = - \sum_i \Pi_i \vec{\nabla} \cdot \vec{v}_i + \text{boundary terms}, \tag{14.17}$$

then $\vec{\nabla} \cdot \vec{v} = 0$ and the boundary conditions imply that energy conservation holds. Thus, Fourier collocation approximation conserves energy when periodic or free-slip boundary conditions are applied to (14.15) [see Fox & Orszag 1973].

15. Survey of Spectral Methods and Applications

In this Section, we give a brief survey of spectral methods and some of their recent applications. There are five important features of spectral methods that should be considered in their formulation and application. They are:

(i) Rate of convergence - If the solution to a problem is infinitely differentiable, then a properly designed spectral method has the property that errors go to zero faster than any finite power of the number of retained modes. In contrast, finite-difference and finite-element methods yield finite-order rates of convergence. The important consequence is that spectral methods can achieve high accuracy with little more resolution than is required to achieve moderate accuracy.

(ii) Efficiency - The development of fast transform methods permits spectral methods to be implemented with comparable efficiency to that of finite difference methods with the same number of independent degrees of freedom. However, since spectral methods typically require a factor of 2-5 fewer degrees of freedom in each space direction to achieve moderate accuracy (say, 5% error), the spectral computations can be considerably more effective. As the required accuracy increases, the attractiveness of spectral methods increases.

(iii) Boundary conditions - As shown in earlier Sections of this monograph, the mathematical features of spectral methods follow very closely those of the partial differential equation being solved. Thus, the boundary conditions imposed on spectral approximations are normally the same as those imposed on the differential equation. In contrast, finite-difference methods of higher order than the differential equation require additional 'boundary conditions.' Many of the complications of finite-order finite-difference methods disappear with the infinite-order-accurate spectral methods.

Another aspect of the treatment of boundary conditions by spectral methods is their high resolution of boundary layers. If the solution to a problem has a boundary layer of thickness ε, then only about $1/\varepsilon^{\frac{1}{2}}$ polynomials [see (3.50)] need be retained to achieve high accuracy. In contrast, finite-difference methods using equally spaced grid points would require about $1/\varepsilon$ grid points to resolve such a boundary layer solution. Moreover, if a coordinate transformation is employed to improve the resolution of a boundary or internal layer of thickness ε, the errors of spectral methods are decreased faster than any finite power of ε as $\varepsilon \to 0$.

(iv) Discontinuities - Surprisingly, spectral methods do a better job of localizing errors than difference schemes and hence require considerably less local dissipation to smooth discontinuities.

(v) Bootstrap estimation of accuracy - It is often possible to estimate the accuracy of spectral computations by examination of the shape of the spectrum. Thus, in computations of three-dimensional incompressible flows at high Reynolds numbers, the mean-square vorticity spectrum must not increase abruptly at large wavenumbers (small scales); if the vorticity spectrum decreases smoothly to 0 as wavenumber increases, it is safe to infer that the calculation is accurate. On the other hand, similar criteria for finite-difference methods can be very misleading.

Let us now survey some applications of spectral methods to incompressible fluid dynamics. We shall classify the method according to the boundary conditions and geometry.

(i) Periodic boundary conditions in Cartesian coordinates - Here Fourier series are appropriate. Spectral methods have been regularly used in three dimensions with 32 x 32 x 32 modes and in two dimensions with 128 x 128 modes to simulate homogeneous turbulence. Most operational codes now use pseudospectral (collocation) methods because aliasing errors are usually small. The key fast transform methods are described in detail by Orszag (1971c).

More recently, more ambitious spectral codes have been developed. The KILOBOX code employs 1024 x 1024 Fourier modes in two dimensions while the CENTICUBE code uses up to 128 x 128 x 128 modes in three dimensions. These high resolution codes are now being used to study fundamental questions regarding high Reynolds number turbulence, including the structure of inertial ranges.

(ii) Rigid boundary conditions in Cartesian coordinates - Here Chebyshev or Legendre polynomials should be employed. Typical applications to date include numerical studies of turbulent shear flows and boundary layer transition. Pseudospectral(collocation) methods are normally used, with Chebyshev polynomials particularly convenient because fast Fourier transform methods can be applied.

(iii) <u>Rigid boundary conditions in cylindrical geometry</u> - Here Chebyshev or Legendre polynomials should be used in radius, Fourier series in angle, and either Fourier or Chebyshev series in the axial direction (depending on boundary conditions). Some technical aspects of the implementation of Chebyshev series in radius, including pole conditions, are discussed by Orszag (1974). Applications to date include studies of transition in circular Couette flow and pipe Poiseuille flow. In particular, it should be emphasized that Chebyshev polynomial expansions are much better suited for serious numerical work than the apparently more natural choice of Bessel function expansions in radius. There are two reasons: Chebyshev series converge faster to general functions regardless of their boundary conditions and Chebyshev-spectral methods can be implemented efficiently by fast transform methods.

(iv) <u>Problems in spherical geometry</u> - Here surface harmonic expansions, generalized Fourier series, and 'associated' Chebyshev expansions all have attractive features. A detailed discussion of these methods is outside the scope of this monograph, but roughly speaking generalized Fourier series permit the most efficient transform methods to be developed followed by associated Chebyshev expansions and then surface harmonic expansions but surface harmonic expansions are best with regard to the pole problem. A variety of applications of these methods to global atmospheric flows have been made.

(v) <u>Semi-infinite or infinite geometry</u> - Here Chebyshev expansions are best if the domain can be mapped or truncated to a finite domain without serious error. There are two cases here: additional boundary conditions may or may not be required at 'infinity.' Here again the formulation of spectral methods follows closely the exact mathematics. If additional boundary conditions, like radiation or outflow boundary conditions, must be imposed on the truncated domain, then they should also be applied to the spectral method. On the other hand, if mapping without additional boundary conditions does not introduce a singularity in the exact equations, no boundary conditions at 'infinity' are required in the spectral approximation.

REFERENCES

Sec. 1: Courant & Hilbert (1953), Jeffreys & Jeffreys (1966), Kantorovich & Krylov (1964).

Sec. 2: Collatz (1960), Fox & Parker (1968), Kantorovich & Krylov (1964), Lanczos (1956), Orszag (1971a,b,c,1972), Richtmyer & Morton (1967), Strang & Fix (1973).

Sec. 3: Acton (1970), Courant & Hilbert (1953), Erdelyi et al (1953), Fox & Parker (1968), Isaacson & Keller (1966), Lanczos (1956), Orszag & Israeli (1974), Rivlin (1969), Szegö (1959), Zygmund (1935).

Sec. 4: Godunov & Ryabenkii (1963), Kreiss (1962), Kreiss & Oliger (1973), Laptev (1975), Miller & Strang (1965), Richtmyer & Morton (1967).

Sec. 5: Barnett & Storey (1974), Bartels & Stewart (1972), Chorin et al (1977), Lie & Engel (1888), Strang (1960), Richtmyer & Morton (1967).

Sec. 6: Fornberg (1975), Kreiss & Oliger (1973), Lanczos (1956, 1966), Lyness (1974), Orszag (1971c,1972), Swartz & Wendroff (1969), Wengle & Seinfeld (1977).

Sec. 7: Price & Varga (1970), Richtmyer & Morton (1967), Swartz & Wendroff (1969).

Sec. 9: Gazdag (1976), Gottlieb & Gustaffson (1976), Kwizak & Robert (1971), Mesinger & Arakawa (1976), Richtmyer & Morton (1967), Roache (1972), Widlund (1966).

Sec. 10: Cooley & Tukey (1965), Cooley (1967), Metcalfe (1974), Orszag (1970, 1971c, 1974), Orszag & Israeli (1974), Patterson & Orszag (1971).

Sec. 11: Boris & Book (1976), Browning et al (1973), Dubiner (1977), Lanczos (1966,1973), Orszag (1971a), Orszag & Israeli (1974), Orszag & Jayne (1974).

Sec. 12: Haidvogel (1977), Roache (1972).

Sec. 13: Deville & Orszag (1977), Gary & Helgason (1970), Orszag (1971b,d,1976a), Orszag & Israeli (1974).

Sec. 14: Buzbee et al (1971), Concus & Golub (1973), Fox & Orszag (1973), Metcalfe (1974), Orszag (1974), Orszag & Israeli (1974), Tang (1977), Zang & Haidvogel (1977).

Sec. 15: Armstrong et al (1970), Bourke (1972), Boyd (1977a,b), Collins & Dennis (1973), Deville & Orszag (1977), Eliasen et al (1970), Fox & Orszag (1973), Francis (1972), Gazdag (1975), Grosch & Orszag (1977), Herbert (1976), Herring et al (1974), Hoskins (1973), Israeli & Orszag (1976), Machenauer (1972), Merilees (1973), Merilees & Orszag (1977), Munson & Joseph (1971), Murdock (1977), Orszag (1970, 1974, 1976a,b,1977), Orszag & Israeli (1974), Orszag & Pao (1974), Orszag & Patterson (1972), Schamel & Elsässer (1976), Tang & Orszag (1977).

BIBLIOGRAPHY

Acton, F. S. 1970 Numerical Methods That Work, Harper & Row, New York.

Armstrong, T. P., Harding, R. C., Knorr, G. & Montogmery, D. 1970 Solution of Vlasov's Equation by Transform Methods, Methods in Computational Physics, Vol, 9, Academic, New York, p. 29

Barnett, S. & Storey, C. 1974 Matrix Methods in Stability Theory, Barnes & Noble, New York.

Bartels, R. & Stewart, G. 1972 Solution of the Matrix Equatión AX + XB = C, Comm. Assoc. Comp. Mach. **15**, 820.

Bourke, W., 1972 An Efficient, One-Level, Primitive-Equation Spectral Model, Mon. Wea. Rev. **100**, 683-689.

Boris, J. P. & Book, D. L. 1976 Flux-Corrected Transport. III- Minimal Error FCT Algorithms, J. Comp. Phys. **20**, 397.

Boyd, J. P. 1977a Pseudospectral Methods for Eigenvalue and Nonseparable Boundary Value Problems, to be published.

Boyd, J. P. 1977b The Choice of Spectral Functions on a Sphere: A Comparison of Chebyshev, Fourier and Associated Legendre Expansions, to be published.

Browning, G., Kreiss, H.O. & Oliger J. 1973 Mesh Refinement, Math. Comp. **27**, 29.

Buzbee, B. L., Dorr, F. W., George, J. A.& Golub, G. H. 1971 The Direct Solution of the Discrete Poisson Equation on Irregular Regions, SIAM J. Num. Anal. **8**, 722.

Chorin, A. J., Hughes, T. J. R., McCracken, M. F. & Marsden, J. E. 1977 Product Formulas and Numerical Algorithms, to be published.

Collatz, L. 1960 The Numerical Treatment of Differential Equations, Springer-Verlag, Berlin.

Collins, W. M. & Dennis, S.C. R. 1973 Flow Past an Impulsively Started Circular Cylinder, J. Fluid Mech. **60**, 105.

Concus P. & Golub, G. H. 1973 Use of Fast Direct Methods for the Efficient Numerical Solution of Nonseparable Elliptic Equations, SIAM J. Num. Anal. **10**, 1103.

Cooley, J. W. 1967 Applications of the Fast Fourier Transform Method, Proc. IBM Scientific Computing Symposium on Digital Simulation of Continuous Systems, IBM Corp., p. 83.

Cooley, J. W. & Tukey, J. W. 1965 An Algorithm for the Machine Computation of Complex Fourier Series, Math. Comp. **19**, 297.

Courant, R. & Hilbert, D. 1953 Methods of Mathematical Physics, Vol. 1, Interscience, New York.

Deville, M. & Orszag, S. A. 1977 To be published.

Dubiner, M. 1977 To be published.

Eliasen, E., Machenauer, B. & Rasmussen, E. 1970 On a Numerical Method for Integration of the Hydrodynamical Equations with a Spectral Representation of the Horizontal Fields. Department of Meterology, Copenhagen University, Denmark, Rep. No. 2, 35 pp.

Erdélyi, A., Magnus, W., Oberhettinger, F. & Tricomi, F.G. 1953 Higher Transcendental Functions, vol. 2, McGraw-Hill, New York.

Fox, D. G. & Orszag, S. A. 1973 Pseudospectral Approximation to Two-Dimensional Turbulence, J. Comp. Phys. 11, 612.

Fox, L. & Parker, I. B. 1968 Chebyshev Polynomials in Numerical Analysis, Oxford University Press, London.

Fornberg, B. 1975 On à Fourier Method for the Integration of Hyperbolic Equations, SIAM J. Num. Anal. 12, 509.

Francis, P. E. 1972 The Possible Use of Laguerre Polynomials for Representing the Vertical Structure of Numerical Models of the Atmosphere, Quart. J. Roy. Met. Soc. 98, 662.

Gary, J. & Helgason, R. 1970 A Matrix Method for Ordinary Differential Equations, J. Comp. Phys. 5, 169.

Gazdag, J. 1975 Numerical Solution of the Vlasov Equation with the Accurate Space Difference Method, J. Comp. Phys. 19, 77.

Gazdag, J. 1976 Time-Differencing Schemes and Transform Methods, J. Comp. Phys. 20, 196.

Godunov, S. K. & Ryabenkii, V. S. 1963 Special Criteria of Stability of Boundary-Value Problems for Non-Self-Adjoint Difference Equations, Uspekhi Mat. Nauk 3, 211.

Gottlieb, D. & Gustaffson, B. 1976 Generalized DuFort-Frankel Methods for Parabolic Initial-Boundary Value Problems, SIAM J. Num. Anal. 13, 129.

Grosch, C. E. & Orszag, S. A. 1977 Numerical Solution of Problems in Unbounded Regions: Coordinate Transforms, J. Comp. Phys., to appear.

Haidvogel, D. 1977 Resolution of Downstream Boundary Layers in the Chebyshev Approximation to Viscous Flow Problems, to be published.

Herbert, T. H. 1976 Periodic Secondary Solutions in a Plane Channel, Proc. Fifth Int'l Conf. on Numerical Methods in Fluid Dynamics (ed. by A.I. van de Vooren and P.J. Zandbergen) Springer-Verlag, Berlin, p.235.

Herring, J. R., Orszag, S. A., Kardman, R. H. & Fox, D. G. 1974 Decay of Two-Dimensional Homogeneous Isotrope Turbulence, J. Fluid Mech. 66, 417.

Hoskins, B. J. Comments on "The Possible Use of Laguerre Polynomials for Representing the Vertical Structure of Numerical Models of the Atmosphere, Quart. J. Roy. Met. Soc. 99, 571.

Isaacson, E. & Keller, H. B. 1966 Analysis of Numerical Methods, Wiley, New York.

Israeli, M. & Orszag, S. A. 1976 Numerical Investigation of Viscous Effects on Trapped Oscillations in a Rotating Fluid, Proc. Fifth Int'l Conf. on Numerical Methods in Fluid Dynamics (ed. by A.I. van de Vooren and P.J. Zandbergen) Springer-Verlag, Berlin, p.241.

Jeffreys, H. & Jeffreys, B. S. 1966 Methods of Mathematical Physics, Cambridge University Press, Cambridge.

Kantorovich, L. V. & Krylov, V. I. 1964 Approximate Methods of Higher Analysis, Noordhoff, Groningen.

Kreiss, H. O. 1962 Über die Stabilitätsdefinition für Differenzengleichungen die partielle Differentialgleichungen approximieren, Nordisk Tidskr. Informations-Behandling 2, 153.

Kreiss, H. O. & Oliger, J. 1973 Methods for the Approximate Solution of Time Dependent Problems, World Meteorological Organization/International Council of Scientific Unions.

Kwizak, M. & Robert, A. J. 1971 A Semi-Implicit Scheme for Grid Point Atmospheric Models of the Primitive Equations, Mon. Weather Rev. 99, 32.

Lanczos, C. 1956 Applied Analysis, Prentice-Hall, Englewood Cliffs, New Jersey.

Lanczos, C. 1966 Discourse on Fourier Series, Hafner, New York.

Lanczos, C. 1973 Legendre versus Chevyshev Polynomials, Topics in Numerical Analysis (ed. by J.J. Miller) Academic, N.Y., p. 191.

Laptev, G. 1975 Conditions for the Uniform Well-Posedness of the Cauchy Problem for Systems of Equations, Soviet Math. Dokl. 16, 65.

Lie, S. & Engel F. 1888 Theorie der Transformationsgruppen, Leipzig.

Lyness, J. 1974 Computational Techniques Based on the Lanczos Representation, Math. Comp. 28, 81.

Machenauer, B. & Rasmussen, E. 1972 On the Integration of the Spectral Hydrodynamical Equations by a Transform Method, Dept. of Meteorology, Copenhagen University, Denmark, Rep. No. 3, 44 pp.

Merilees, P.E. 1973 Pseudo-Spectral Approximation Applied to the Shallow Water Equations on a Sphere, Atmosphere 11, 13.

Merilees, P. E. & Orszag, S. A. 1977 To be published.

Mesinger & Arakawa, A. 1976 Numerical Methods Used in Atmospheric Models, Vol. 1, World Meteorological Organization/International Council of Scientific Unions.

Metcalfe R. W. 1974 Spectral Methods for Boundary Value Problems in Fluid Mechanics, Ph. D. Thesis, M.I.T., Cambridge, Mass.

Miller, J. J. H. & Strang, W. G. 1965 Matrix Theorems for Partial Differential and Difference Equations, Stanford University Tech. Report CS28, Stanford, California.

Munson, B. R., & D. D. Joseph 1971 Viscous Incompressible Flow Between Concentric Rotating Spheres, Part 1, Basic Flow, J. Fluid Mech.,49 , 289.

Murdock, J. W. 1977 A Numerical Study of Nonlinear Effects on Boundary Layer Stability, AIAA Paper 77-127.

Orszag, S. A. 1970 Transform Method for the Calculation of Vector-Coupled Sums: Application to the Spectral Form of the Vorticity Equation, J. Atmos. Sci. 27, 890.

Orszag, S. A. 1971a Numerical Simulation of Incompressible Flows within Simple Boundaries: Accuracy, J. Fluid Mech. 49, 75.

Orszag, S. A. 1971b Galerkin Approximations to Flows with Slabs, Spheres, and Cylinders, Phys. Rev. Letters 26, 1100 (1971).

Orszag, S. A. 1971c Numerical Simulation of Incompressible Flows within Simple Boundaries: Galerkin (Spectral) Representations, Stud. in Appl. Math. 50, 395.

Orszag, S. A. 1971d Accurate Solution of the Orr-Sommerfeld Equation, J. Fluid Mech. 50, 689.

Orszag, S. A. 1972 Comparison of Pseudospectral and Spectral Approximation, Stud. in Appl. Math. 51, 253.

Orszag, S. A. 1974 Fourier Series on Spheres, Mon. Weather Rev. 102, 56.

Orszag, S. A. 1976a Turbulence and Transition: A Progress Report, Proc. Fifth Int'l Conf. on Numerical Methods in Fluid Dynamics (ed. by A.I. van de Vooren and P. J. Zandbergen), Springer-Verlag, Berlin, p.32.

Orszag, S. A. 1976b Design of Large Hydrodynamics Codes, Computer Science and Scientific Computing (ed. by J. Ortega), Academic, New York, p. 191.

Orszag, S. A. 1977 Numerical Simulation of Turbulent Flows, Handbook of Turbulence, Plenum, New York, p. 281.

Orszag, S. A. & Israeli, M. 1974 Numerical Simulation of Viscous Incompressible Flows, Ann. Rev. Fluid Mech. 5, 281.

Orszag, S. A. & Jayne, L. W. 1974 Local Errors of Approximate Solutions to Hyperbolic Problems, J. Comp. Phys. 14, 93.

Orszag, S.A. & Pao, Y.H. 1974 Numerical Computation of Turbulent Shear Flows, Advances in Geophysics, Vol. 18A (ed. by F.N. Frenkiel and R.E. Munn), Academic, New York, p. 225.

Orszag, S.A. & Patterson, G.S. 1972 Numerical Simulation of Three-Dimensional Homogeneous Isotropic Turbulence, Phys. Rev. Letters 28, 76.

Patterson, G.S. & Orszag, S.A. 1971 Spectral Calculations of Isotropic Turbulence: Efficient Removal of Aliasing Interactions, Phys. Fluids 14, 2538.

Price, H. S. & Varga, R. S. 1970 Error Bounds for Semidiscrete Galerkin Approximations of Parabolic Problems with Applications to Petroleum Reservoir Mechanics, Numerical Solution of Field Problems in Continuum Physics, American Mathematical Society, Providence, p.74.

Richtmyer, R. D. & Morton, K. W. 1967 Difference Methods for Initial Value Problems, Interscience, New York.

Rivlin, T. J. 1969 An Introduction to the Approximation of Functions, Blaisdell, Waltham, Mass.

Roache, P. J. 1972 Computational Fluid Dynamics, Hermosa Publishers, Albuquerque, N. M.

Schamel, H. & Elsasser, K. 1976 The Application of the Spectral Method to Nonlinear Wave Propagation, J. Comp. Phys. 22, 501.

Strang, W. G. 1960 Difference Methods for Mixed Boundary-Value Problems, Duke Math. J. 27, 221.

Strang, G. & Fix, G. J. 1973 An Analysis of the Finite Element Method, Prentice-Hall, Englewood Cliffs, New Jersey.

Swartz, B. K. & Wendroff, B. 1969 Generalized Finite Difference Schemes, Math. Comp. 23, 37.

Szegö, G. 1959 Orthogonal Polynomials, American Mathematical Society, New York.

Tang(Hui), C. M. 1977 Numerical Simulation of Fluid Flow in Spherical and Two-Dimensional Magnetohydrodynamic Turbulence, Ph.D. Thesis, M. I. T., Cambridge, Mass.

Tang, C. M. & Orszag, S. A. 1977 Two-Dimensional Turbulence on the Surface of a Sphere, to be published.

Wengle, H. & Seinfeld, J. 1977 Pseudo-Spectral Solution of Atmospheric Diffusion Problems, J. Comp. Phys., to appear.

Widlund, O. 1966 Stability of Parabolic Difference Schemes in the Maximum Norm, Numer. Math. 8, 186.

Zang, T. & Haidvogel, D. 1977 To be published.

Zygmund, A. 1935, Trigonometrical Series, Dover, New York.

Appendix. Properties of Chebyshev Polynomial Expansions

The Chebyshev polynomial of degree n, $T_n(x)$, is defined by

$$T_n(\cos\theta) = \cos n\theta. \tag{A.1}$$

Thus, $T_0(x) - 1$, $T_1(x) = x$, $T_2(x) = 2x^2-1$, $T_3(x) = 4x^3-3x$, $T_4(x) = 8x^4-8x+1$, and so on. The Chebyshev polynomials are the solutions of the differential equation

$$\sqrt{1-x^2}\,\frac{d}{dx}\sqrt{1-x^2}\,\frac{dT_n}{dx} + n^2\,T_n = 0 \tag{A.2}$$

that are bounded at $x = \pm 1$. They satisfy the orthogonality relation

$$\int_{-1}^{1} T_n(x)\ T_m(x)\ (1-x^2)^{-1/2}\,dx = \frac{\pi}{2}\,c_n\,\delta_{nm}, \tag{A.3}$$

where $c_0 = 2$, $c_n = 1$ for $n>0$. Some properties of Chebyshev polynomials are

$$T_{n+1}(x) = 2x\,T_n(x) - T_{n-1}(x), \tag{A.4}$$

$$|T_n(x)| \leq 1, \qquad |T_n'(x)| \leq n^2, \tag{A.5}$$

$$\frac{d^p}{dx^p}\,T_n(\pm 1) = (\pm 1)^{n+p}\prod_{k=0}^{p-1}(n^2-k^2)/(2k+1), \tag{A.6}$$

$$\left|\frac{d^p}{dx^p}\,T_n(x)\right| = 0(n^{2p}) \quad (n\to\infty \quad p \text{ fixed}, |x|\leq 1), \tag{A.7}$$

$$T_n(\pm 1) = (\pm 1)^n,\ T_{2n}(0) = (-1)^n,\ T_{2n+1}(0) = 0, \tag{A.8}$$

$$T_{2n}'(0) = 0, \quad T_{2n+1}'(0) = (-1)^n n.$$

The following formulae relate the expansion coefficients a_n in the series

$$f(x) = \sum_{n=0}^{\infty} a_n T_n(x) \qquad (|x| \le 1)$$

to the expansion coefficients b_n of

$$Lf(x) = \sum_{n=0}^{\infty} b_n T_n(x) \qquad (|x| \le 1)$$

for various linear operators L. We use the constants c_n and d_n defined by

$$c_0 = 2, \ c_n = 0 \quad (n<0), \ c_n = 1 \quad (n>0),$$

$$d_n = 1 \quad (n \ge 0), \quad d_n = 0 \quad (n<0).$$

Some formulae are:

$$Lf = f'(x): \ c_n b_n = 2 \sum_{\substack{p=n+1 \\ p+n \text{ odd}}}^{\infty} p a_p \tag{A.9}$$

$$Lf = f''(x): \ c_n b_n = \sum_{\substack{p=n+2 \\ p+n \text{ even}}}^{\infty} p(p^2-n^2) a_p \tag{A.10}$$

$$Lf = xf(x): \ b_n = \tfrac{1}{2}(c_{n-1}a_{n-1}+a_{n+1}) \tag{A.11}$$

$$Lf = x^2 f(x): \ b_n = \frac{1}{4}\{c_{n-2}a_{n-2}+(c_n+c_{n-1})a_n+a_{n+2}\} \tag{A.12}$$

$$Lf = x^4 f(x) \cdot \quad b_n = \frac{1}{16}\{c_{n-4}a_{n-4} + (c_{n-3} + c_{n-2}^2 + 2c_{n-2})a_{n-2} \tag{A.13}$$

$$+ (c_{n-2} + 2c_{n-1} + c_{n-1}^2 + c_n^2 + c_n)a_n + (c_{n-1} + c_n + c_{n+1} + c_{n+2})a_{n+2} + a_{n+4}\}$$

$$Lf = \frac{f(x)-f(0)}{x} : \quad c_n b_n = 2 \sum_{\substack{p=n+1 \\ p+n \text{ odd}}}^{\infty} (-1)^{\frac{p-n-1}{2}} a_p \tag{A.14}$$

$$Lf = \frac{f(x)-f(0)-f'(0)x}{x^2} : c_n b_n = 2 \sum_{\substack{p=n+2 \\ p+n \text{ even}}}^{\infty} (p-n)(-1)^{\frac{p-n-2}{2}} a_p \tag{A.15}$$

$$Lf = \frac{f'(x)-f'(0)}{x} : \quad c_n b_n = 4 \sum_{\substack{p=n+2 \\ p-n \equiv 2 \bmod 4}}^{\infty} p a_p \tag{A.16}$$

$$Lf = \frac{f'(x)-f'(0)-f''(0)x}{x^2} : \quad c_n b_n = 2\left\{ \sum_{\substack{p=n+3 \\ p-n \equiv 3 \bmod 4}}^{\infty} (p-n+1) p a_p \right. \tag{A.17}$$

$$\left. - \sum_{\substack{p=n+5 \\ p-n \equiv 1 \bmod 4}}^{\infty} (p-n-1) p a_p \right\}$$

$$Lf = xf'(x) : \quad c_n b_n = n a_n + 2 \sum_{\substack{p=n+2 \\ p+n \text{ even}}}^{\infty} p a_p \tag{A.18}$$

$$Lf = x^2 f'(x) : \quad b_n = \frac{1}{2}\{(n-1)a_{n-1} + (n+1)(1+d_{n-1}+c_{n-1})a_{n+1} \tag{A.19}$$

$$+ 4 \sum_{\substack{p=n+3 \\ p+n \text{ odd}}}^{\infty} p a_p\}$$

$$Lf = xf''(x) : c_n b_n = 2n(n+1)a_{n+1} + \sum_{\substack{p=n+3 \\ p+n \text{ odd}}}^{\infty} p(p^2-n^2-1)a_p \qquad (A.20)$$

$$Lf = x^2 f''(x) : c_n b_n = n(n-1)a_n + \sum_{\substack{p=n+2 \\ p+n \text{ even}}}^{\infty} p(p^2-n^2-2)a_p \qquad (A.21)$$

$$Lf = \frac{f(x)}{1-x^2}$$

$$\text{with } f(\pm 1)=0 : c_n b_n = -2 \sum_{\substack{p=n+2 \\ p+n \text{ even}}}^{\infty} (p-n)a_p \qquad (A.22)$$

Also, if we expand $f^{(q)}(x)$ as in

$$\frac{d^q}{dx^q} f(x) = \sum_{n=0}^{\infty} a_n^{(q)} T_n(x),$$

then

$$c_{n-1}a_{n-1}^{(q)} - a_{n+1}^{(q)} = 2na_n^{(q-1)} . \qquad (A.23)$$

Properties of Legendre Polynomial Expansions

The Legendre polynomial of degree n, $P_n(x)$, is defined as the solution of the differential equation

$$\frac{d}{dx}(1-x^2)\frac{dP_n(x)}{dx} + n(n+1)P_n(x) = 0 \tag{A.24}$$

that satisfies $P_n(1) = 1$. Thus, $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = \frac{1}{2}(3x^2 - 1)$, and so on. The Legendre polynomials satisfy the orthogonality relation

$$\int_{-1}^{1} P_n(x)\, P_m(x)\, dx = \frac{2}{2n+1}\, \delta_{nm}. \tag{A.25}$$

Some properties of Legendre polynomials are

$$(n+1)\, P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x) \tag{A.26}$$

$$P_n(\pm 1) = (\pm 1)^n,\quad P_n'(\pm 1) = (\pm 1)^{n-1}\, \frac{1}{2}n(n+1) \tag{A.27}$$

$$|P_n(x)| \leq 1,\quad |P_n'(x)| \leq \frac{1}{2}n(n+1) \qquad (|x| \leq 1). \tag{A.28}$$

$$P_n(\cos\theta) = \left[\frac{2}{n\pi\sin\theta}\right]^{1/2} \sin\left[\left(n+\frac{1}{2}\right)\theta + \frac{\pi}{4}\right] + 0(n^{-3/2})$$

$$[n\to\infty;\ \theta\neq\theta,\ \pi;\ \theta \text{ fixed}] \tag{A.29}$$

If $f(x)$ is expanded in the Legendre series

$$f(x) = \sum_{n=0}^{\infty} a_n P_n(x)$$

then

$$a_n = \frac{2}{2n+1} \int_{-1}^{1} f(x) P_n(x) \, dx.$$

If L is a linear operator and

$$Lf(x) = \sum_{n=0}^{\infty} b_n P_n(x)$$

then the relation between b_n and a_n is as follows:

$$Lf(x) - f'(x): \quad b_n = (2n+1) \sum_{\substack{p=n+1 \\ p+n \text{ odd}}}^{\infty} a_p \tag{A.30}$$

$$Lf(x) = f''(x): \quad b_n = (n+\tfrac{1}{2}) \sum_{\substack{p=n+2 \\ p+n \text{ even}}}^{\infty} [p(p+1)-n(n+1)]a_p \tag{A.31}$$

$$Lf(x) = xf(x): \quad b_n = \frac{n}{2n-1} a_{n-1} + \frac{n+1}{2n+3} a_{n+1}. \tag{A.32}$$